

Hybrid On-chip Data Networks

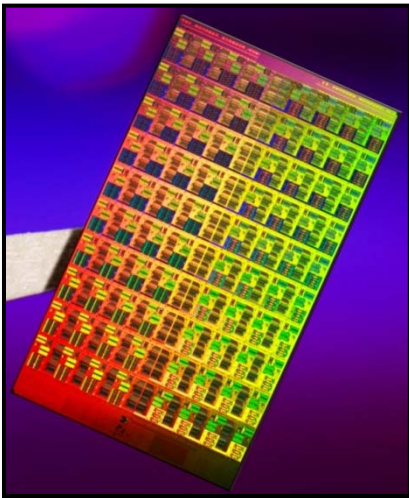
Gilbert Hendry

Keren Bergman

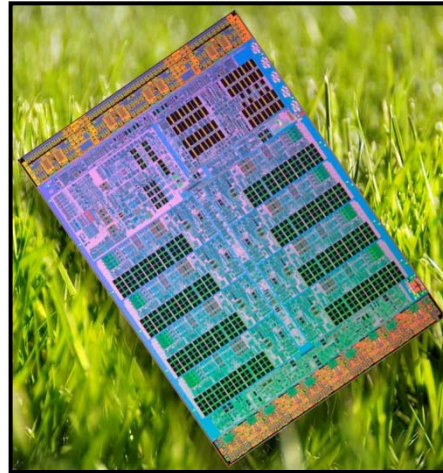


Lightwave Research Lab
Columbia University

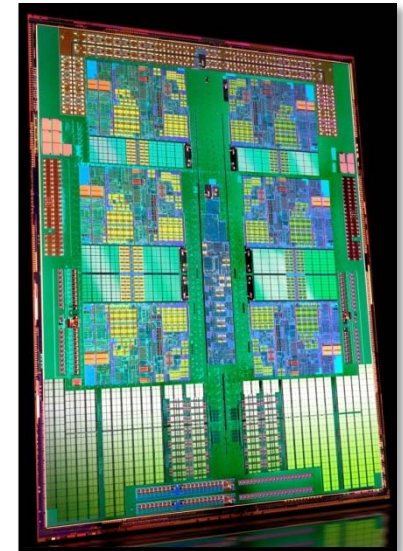
- Chip multi-processors create need for high performance interconnects
- Performance bottleneck of on-chip networks and I/O
- Power dissipation constraints of the chip package
 - > 50% of total power comes from interconnects*



Intel Polaris



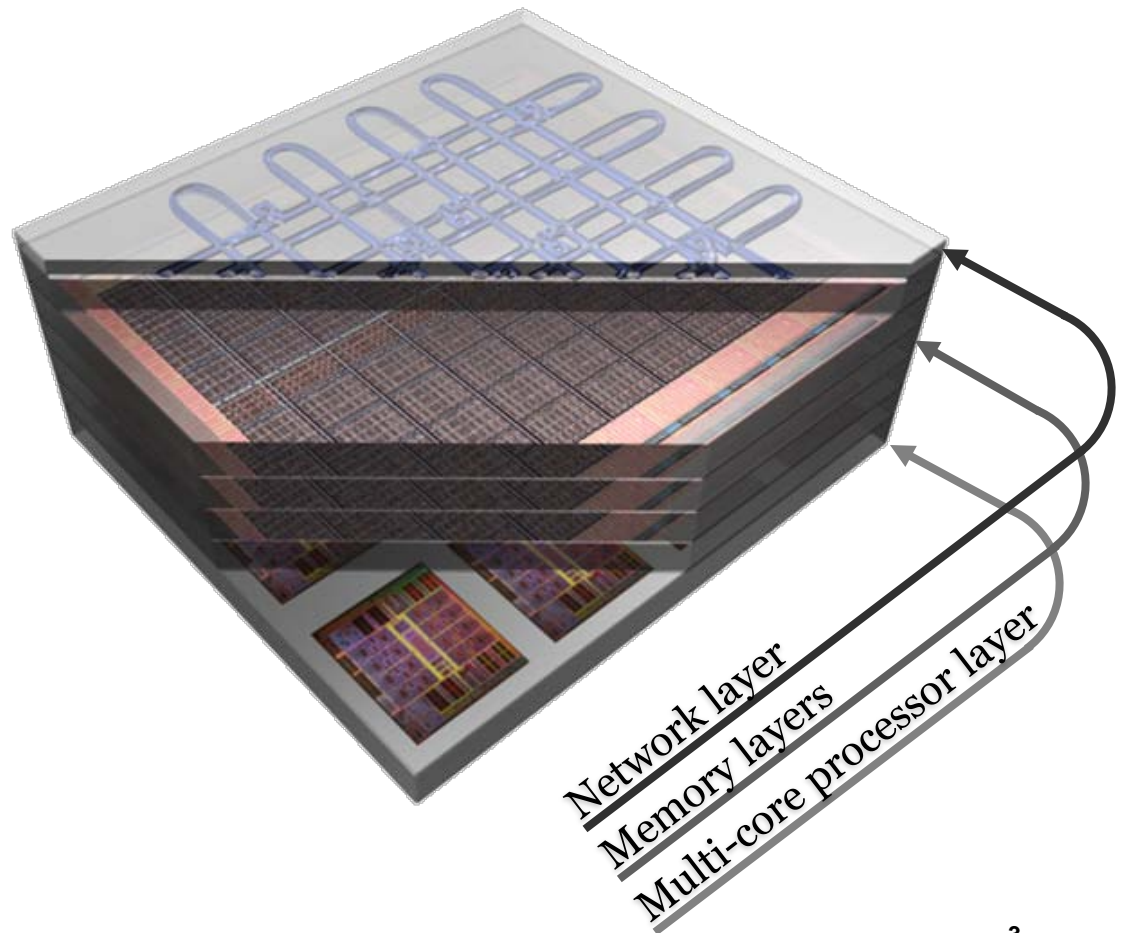
IBM Cell



AMD Opteron

* N. Magen *et al.*, "Interconnect-power dissipation in a microprocessor," SLIP 2004.

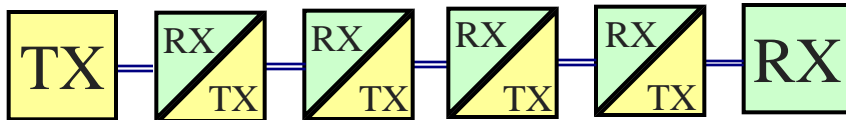
- **CMPs of the future = 3D stacking**
- **Lots of data on chip**
- **Photonics offers key advantages**



Photonics changes the rules for Bandwidth, Energy, and Distance.

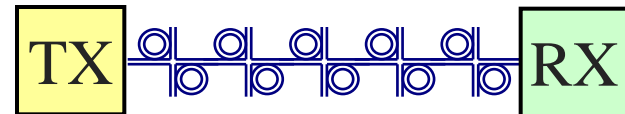
ELECTRONICS:

- Buffer, receive and re-transmit at every router.
- Each bus lane routed independently.
($P \propto N_{\text{LANES}}$)
- Off-chip BW is pin-limited and power hungry.

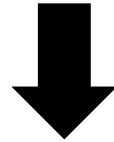


OPTICS:

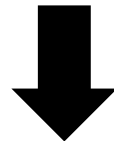
- Modulate/receive high bandwidth data stream once per communication event.
- Broadband switch routes entire multi-wavelength stream.
- Off-chip BW = On-chip BW for nearly same power.



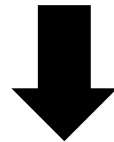
Optical processing difficult and limited



Source, destination routing inefficient



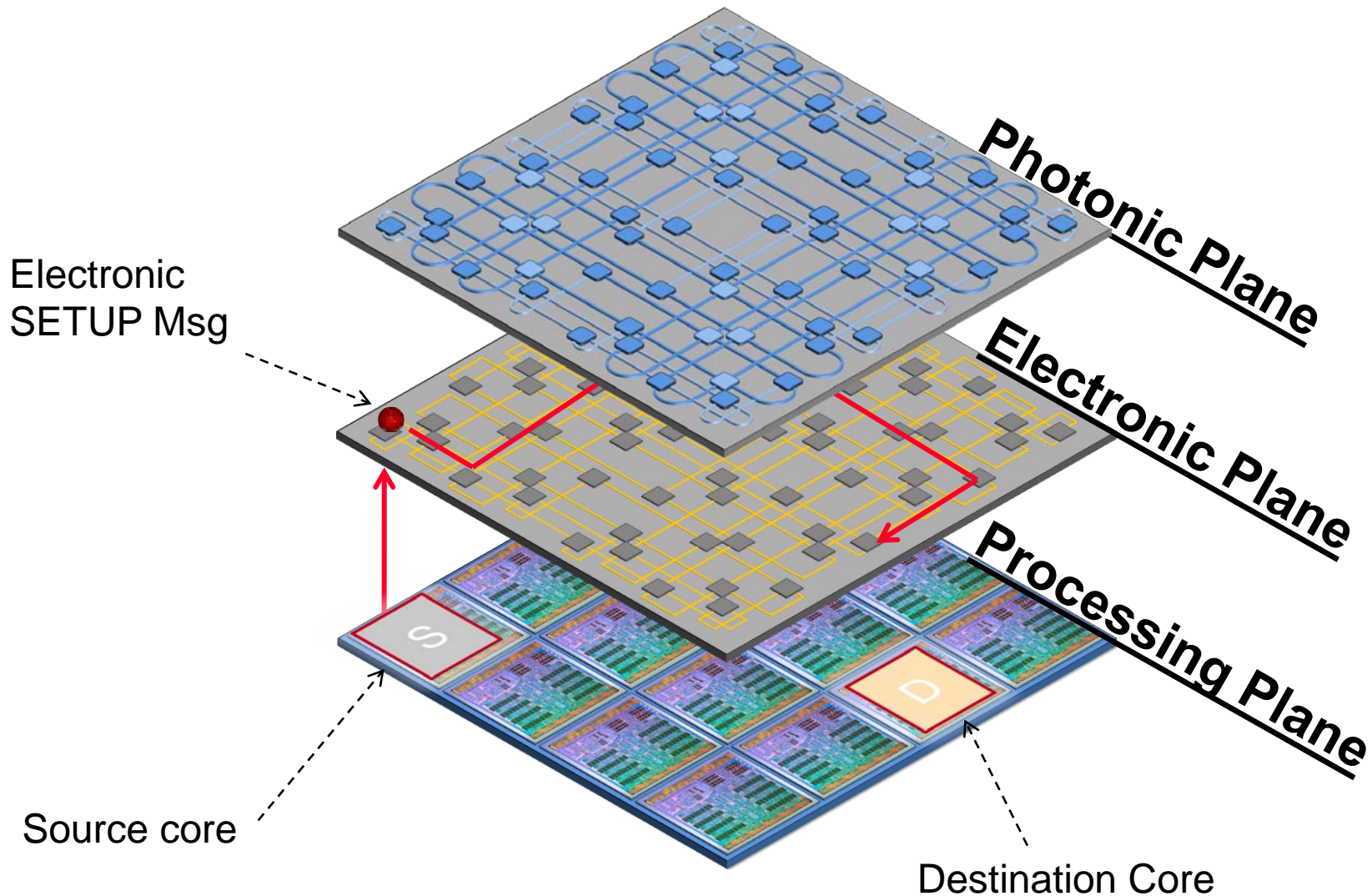
Use electronics for routing,
optics for switching and transmission



Hybrid Circuit-Switching

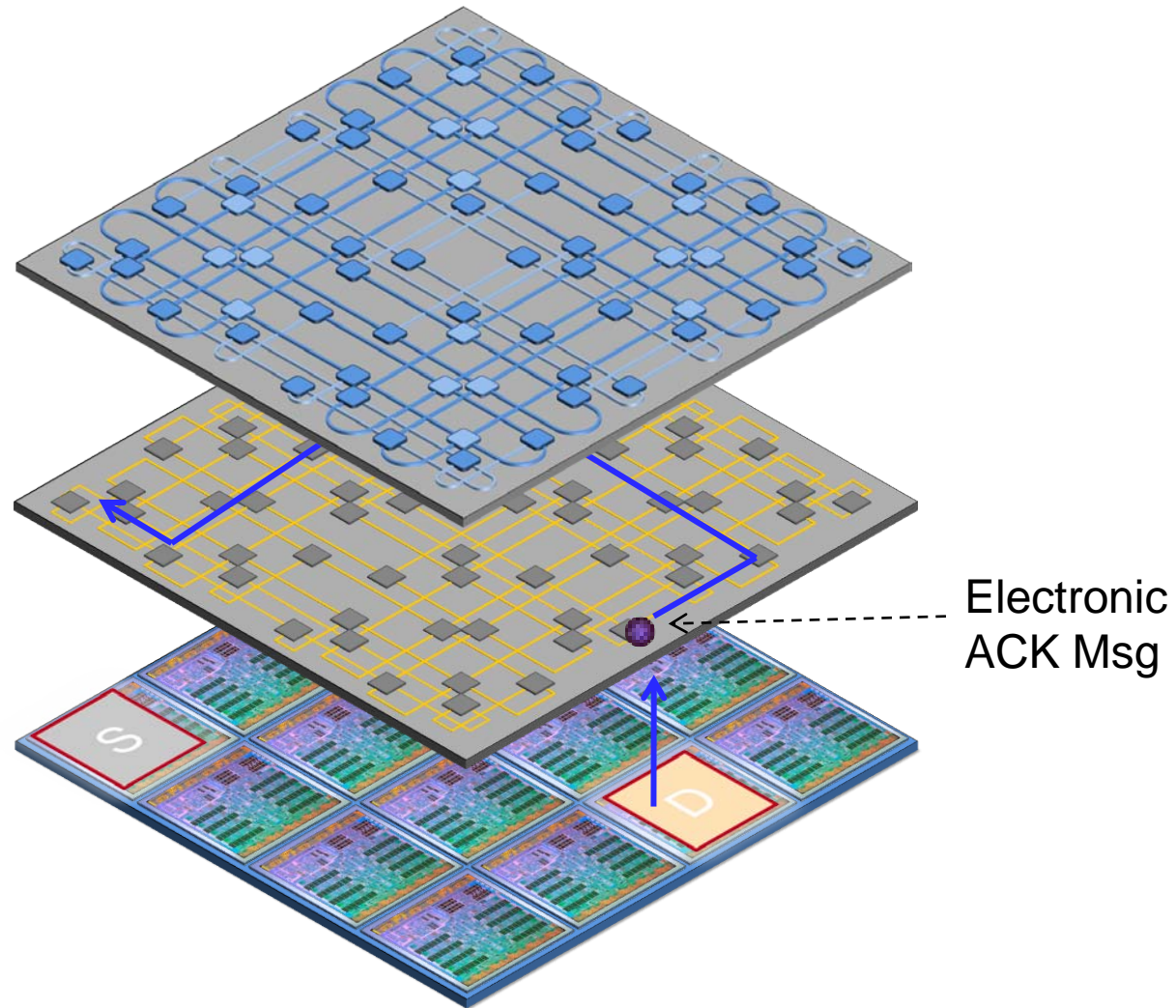
Hybrid Circuit-Switched Networks

Step 1: Path SETUP request



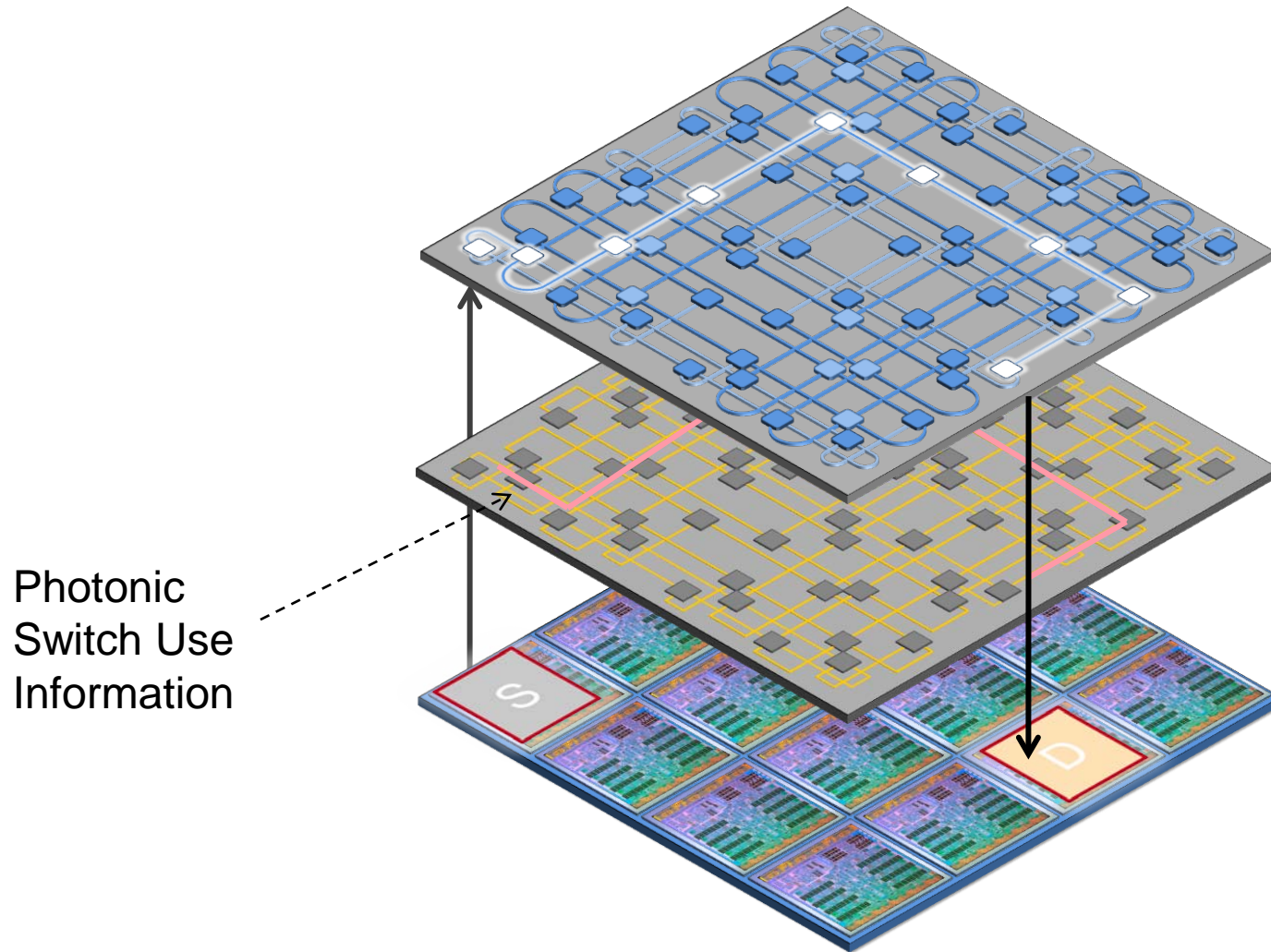
Hybrid Circuit-Switched Networks

Step 2: Path ACK

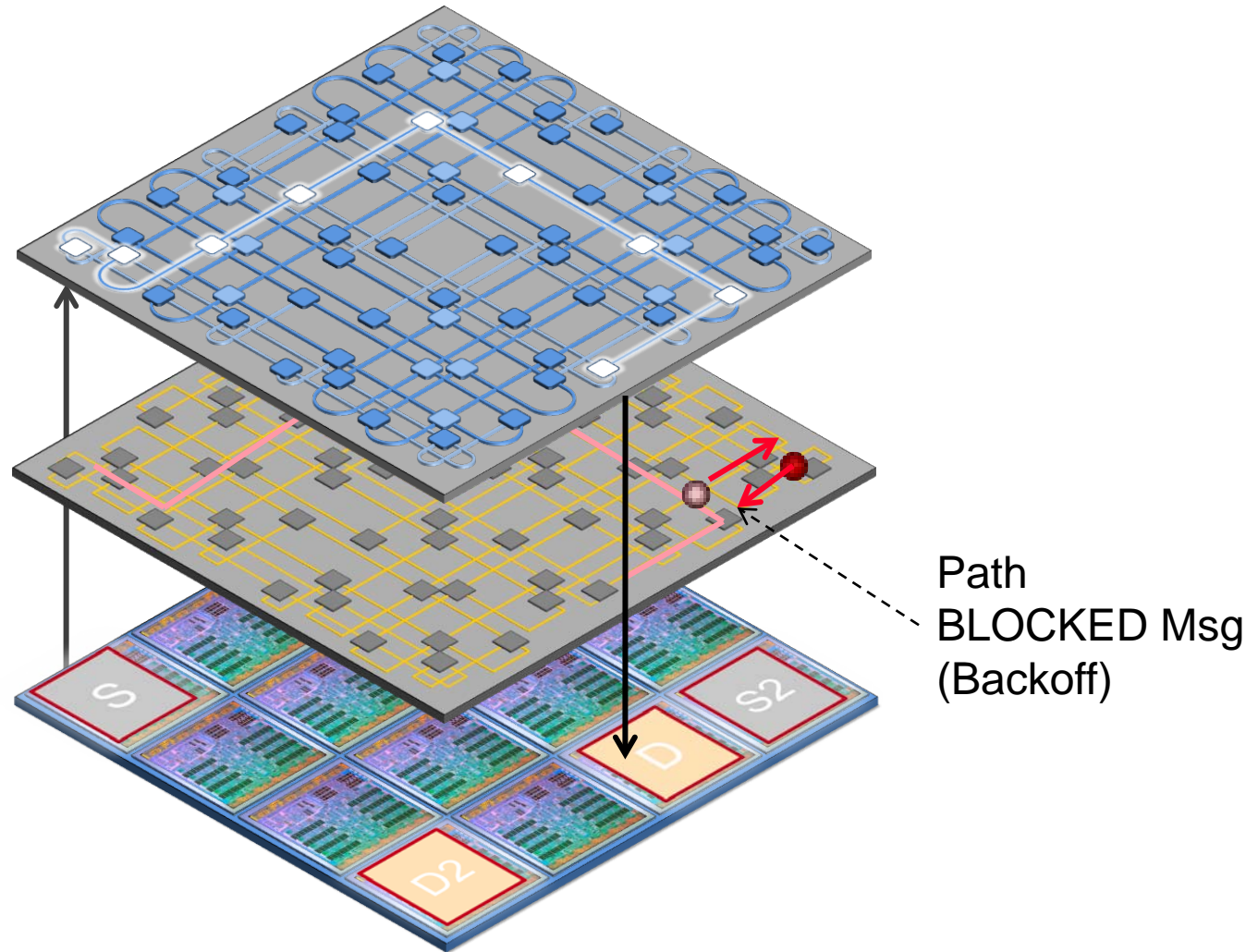


Hybrid Circuit-Switched Networks

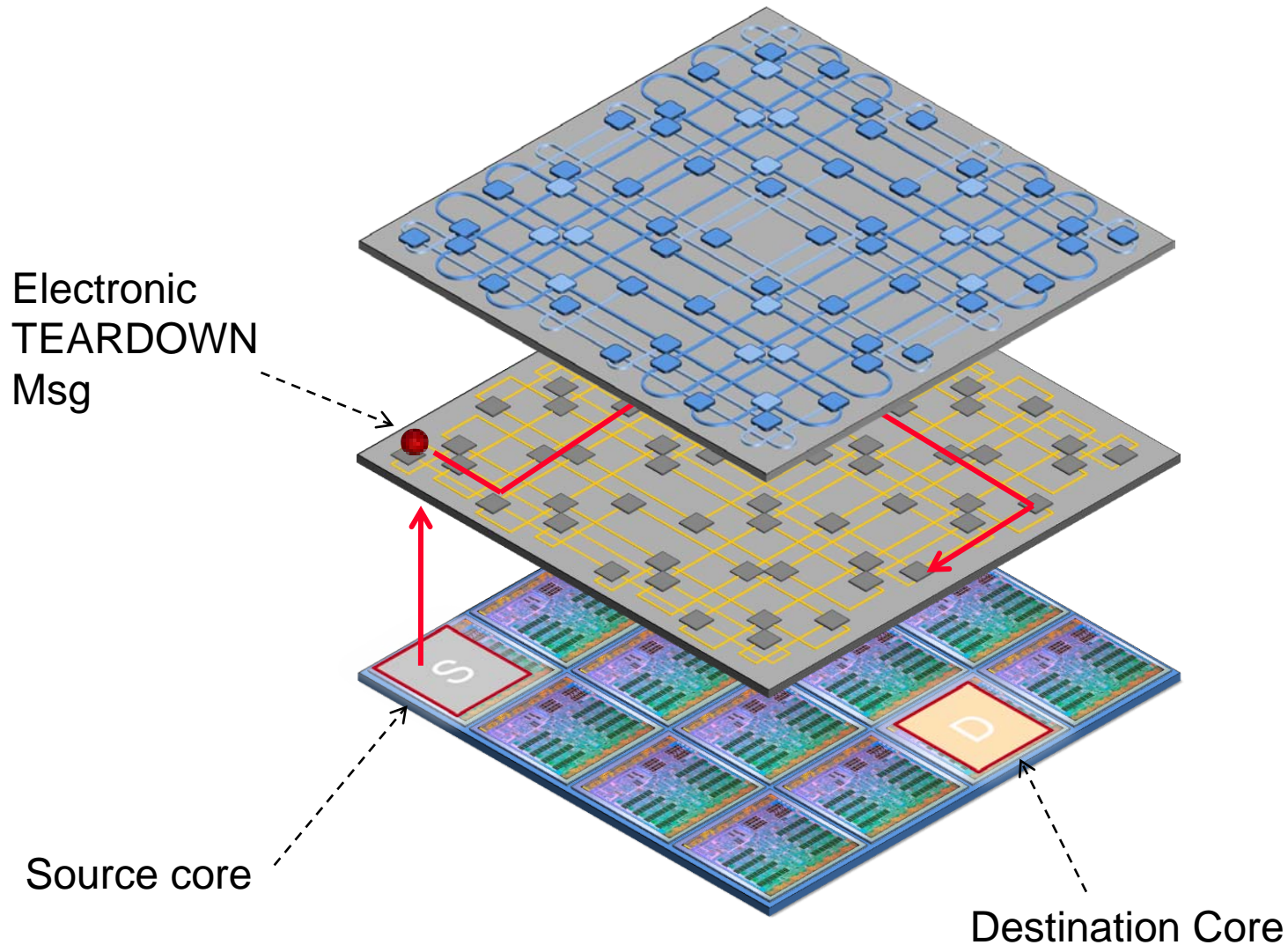
Step 3: Transmit Data



Meanwhile: Path Contention



Step 4: Path TEARDOWN



Pros:

- **Energy-efficient end-to-end transmission**
- **High bandwidth through WDM**
- **Electronic network still available for small control messages***
- **Network-level support for secure regions**

Cons:

- **Path setup latency**
- **Path setup contention (no fairness)**

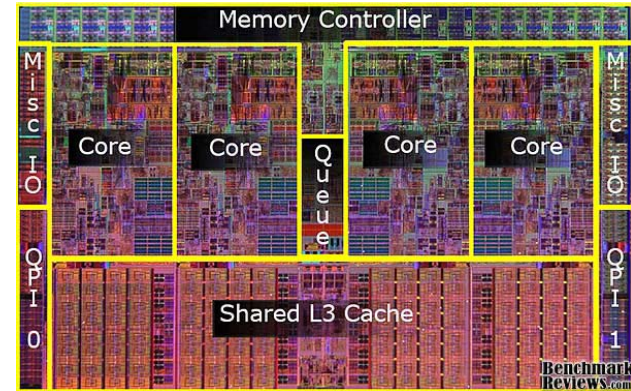
* [G. Hendry et al. *Analysis of Photonic Networks for a Chip Multiprocessor Using Scientific Applications*. In NOCS, 2009]

Programming and Communication

Shared Memory

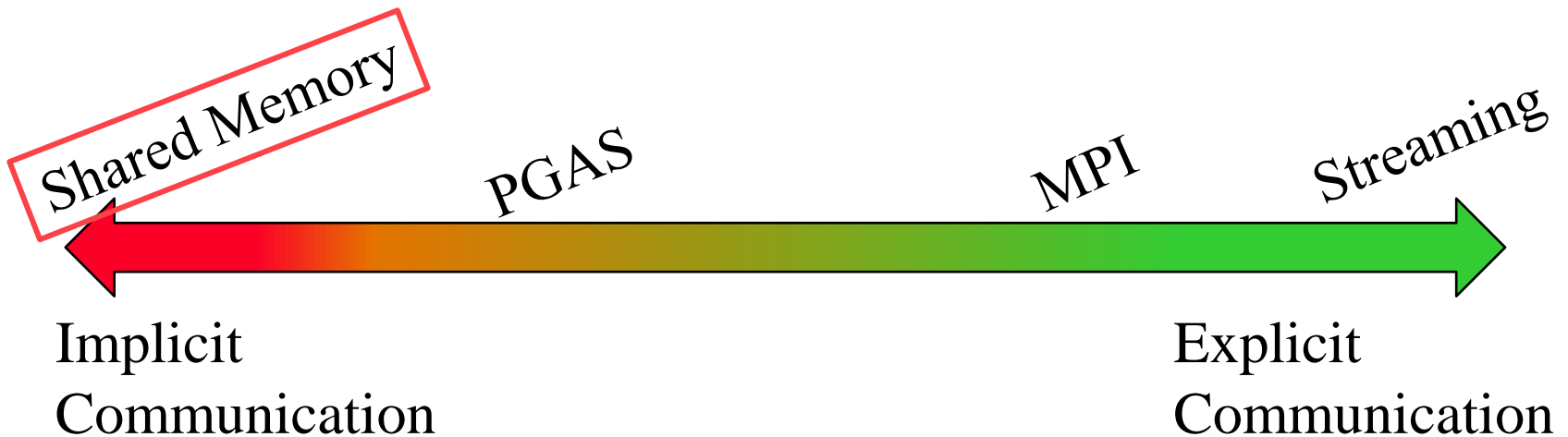


scaling



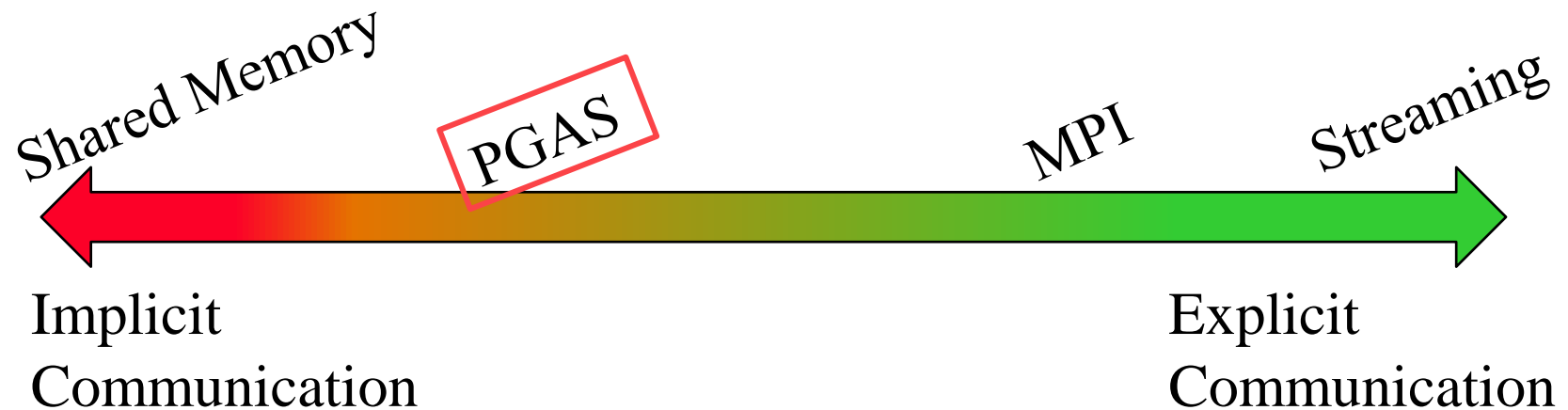
“... [OpenMP on large systems] often performs worse than message passing due to a combination of false sharing, coherence traffic, contention, and system issues that arise from the difference in scheduling and network interface moderation”

~ Exascale Report



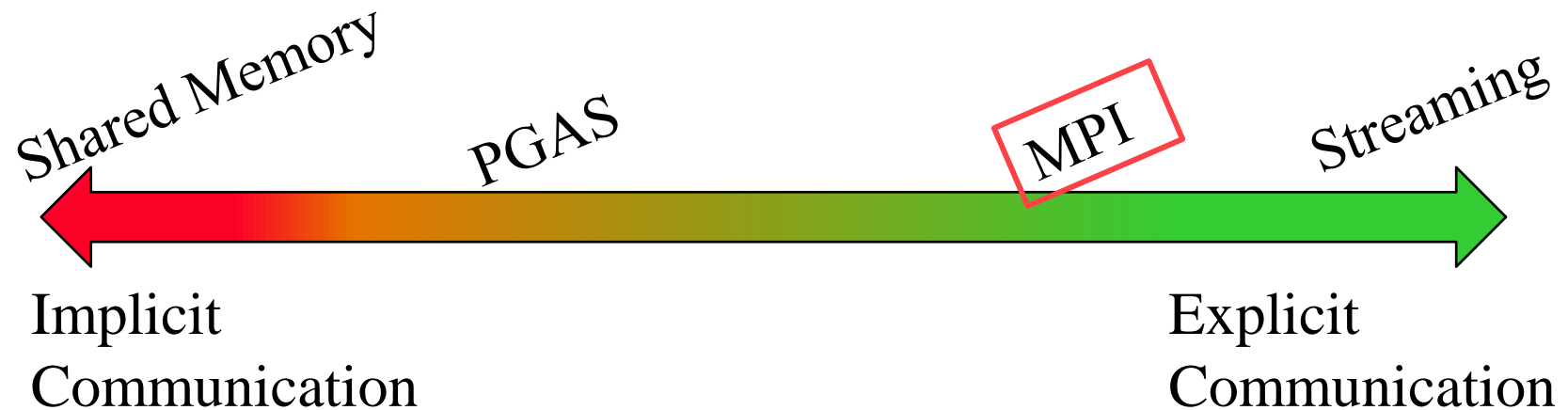
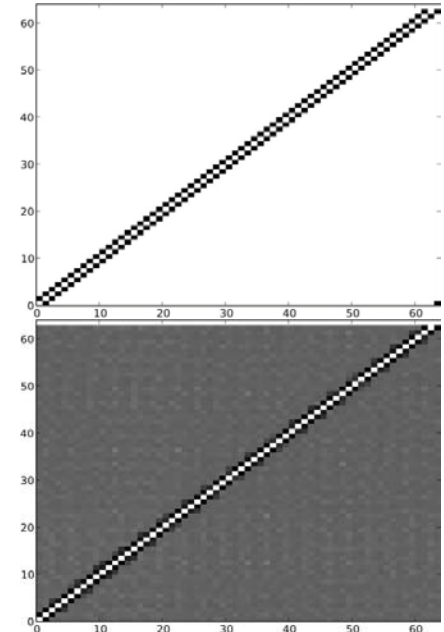
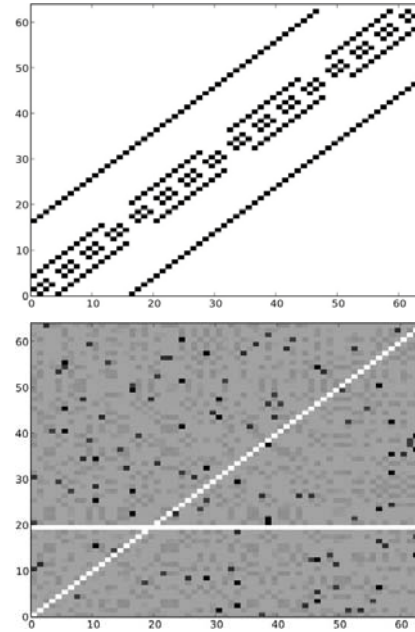
Partitioned Global Address Space

Access	Method
Local Read	Optical Receive
Local Write	Optical send
Remote Read	Electronic request, optical receive
Remote Write	Optical send
Shared R/W	?

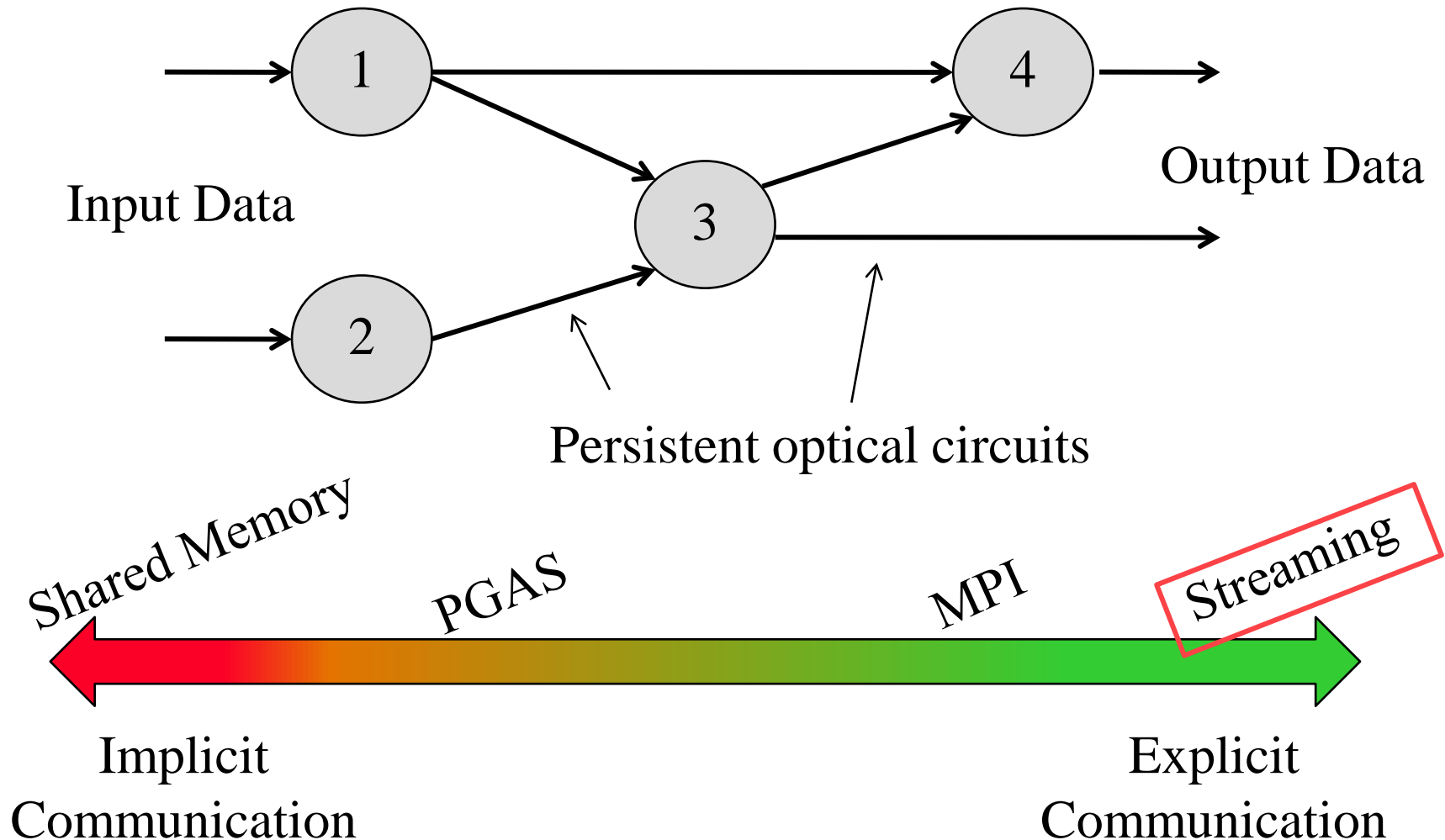


Message Passing

- Complex, dynamic access patterns
- Relatively larger blocks of data
- Scientific computing →

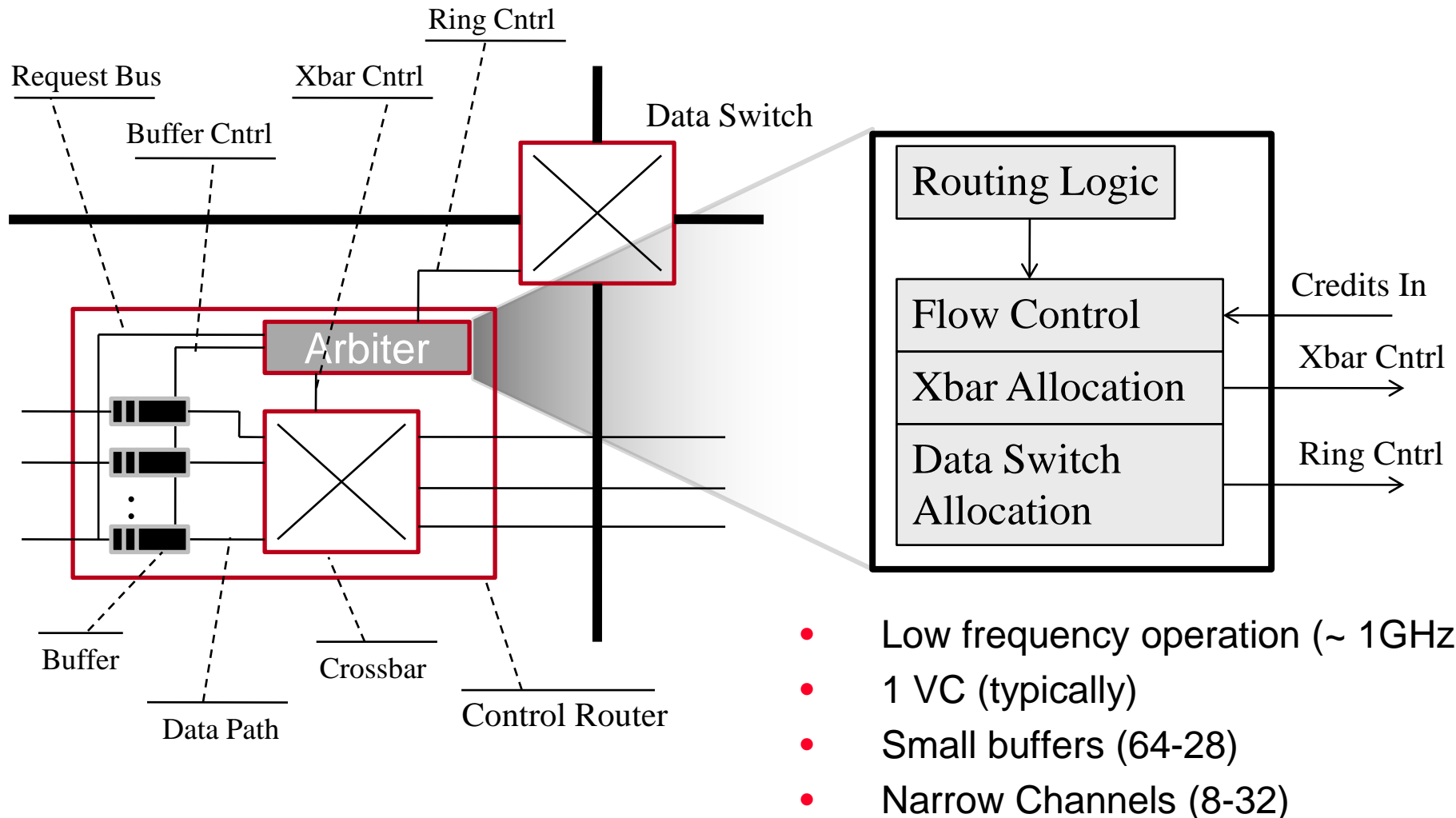


- Embedded / specialized systems (Graphics, Image + Signal Proc.)
- Execution mode of general-purpose systems (Cell Processor)

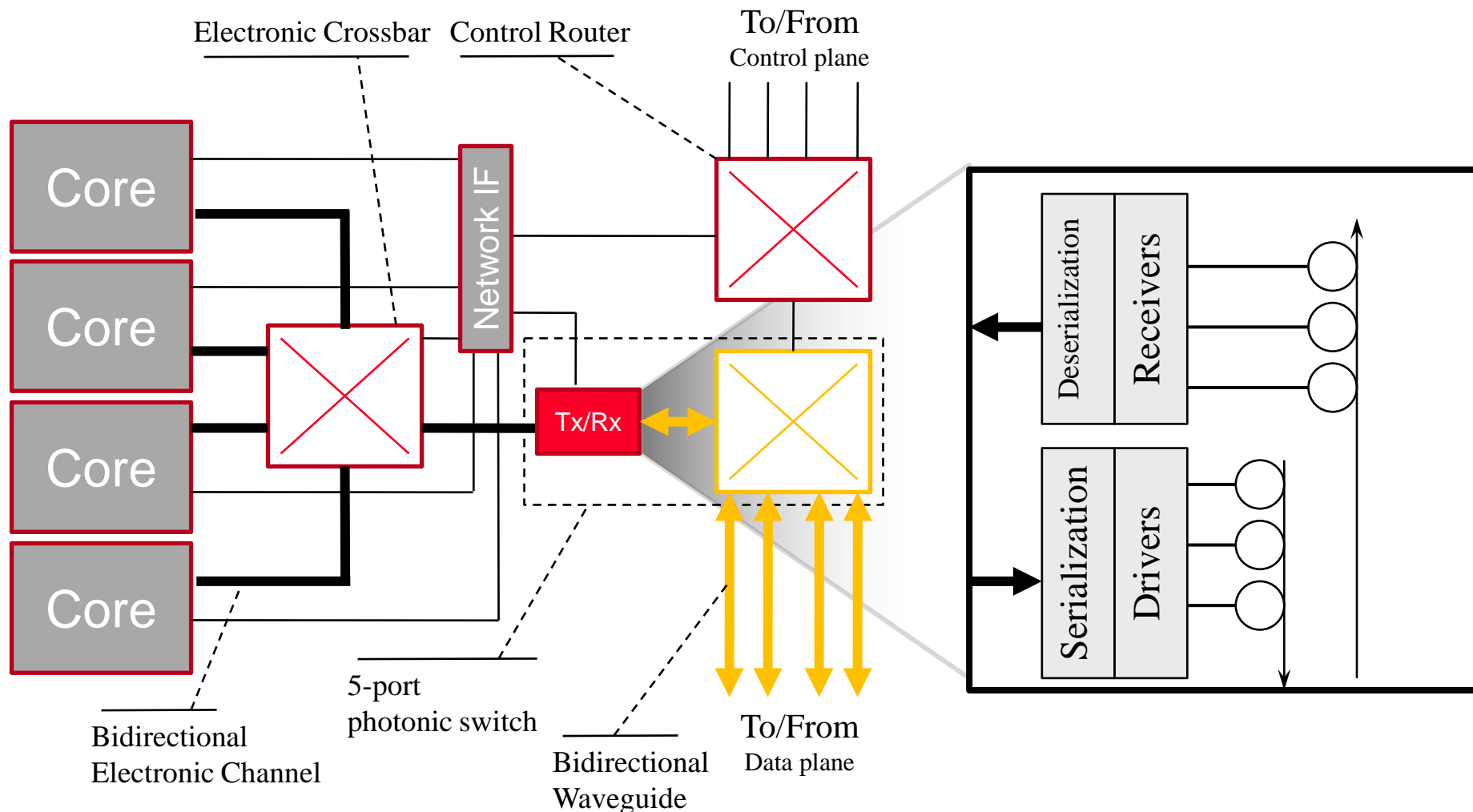


Electronic Plane

Electronic Router



- Low frequency operation ($\sim 1\text{GHz}$)
- 1 VC (typically)
- Small buffers (64-28)
- Narrow Channels (8-32)

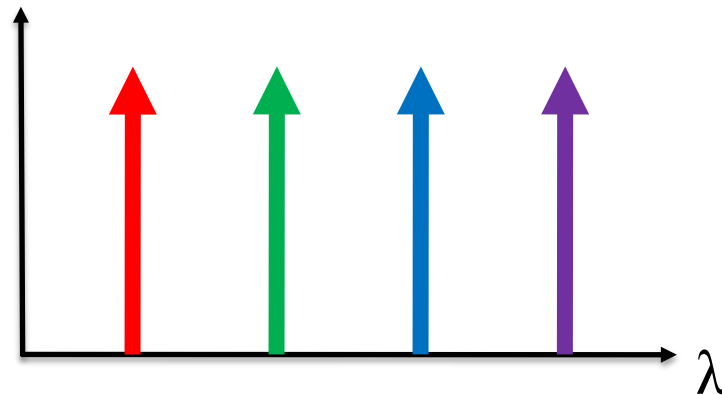
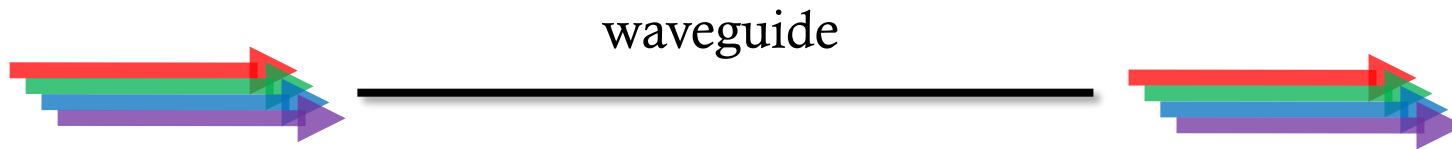


External Concentration

[P. Kumar et al. *Exploring concentration and channel slicing in on-chip network router*. In NOCS, 2009]

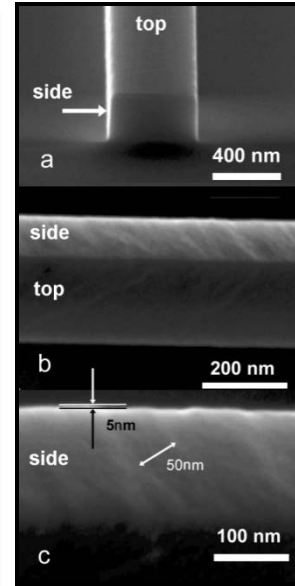
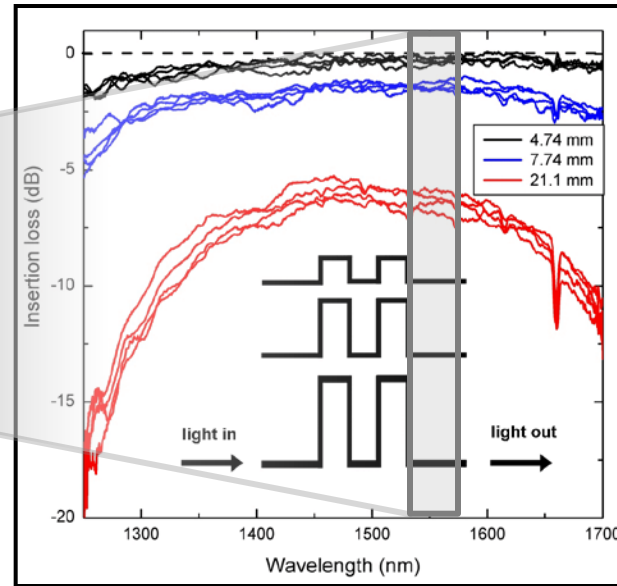
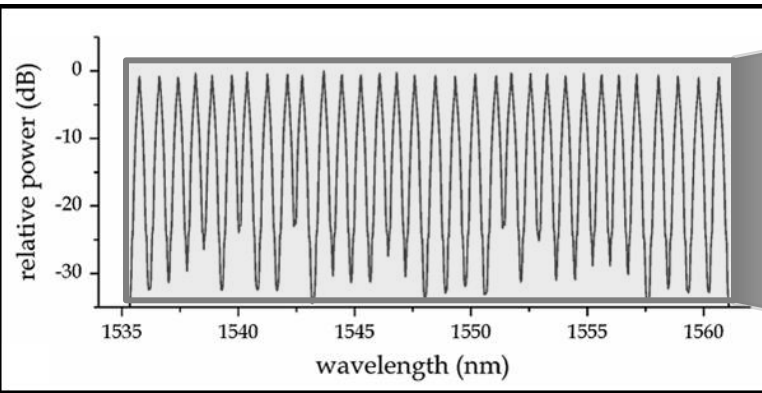
The Photonic Plane

Wavelength Division Multiplexing

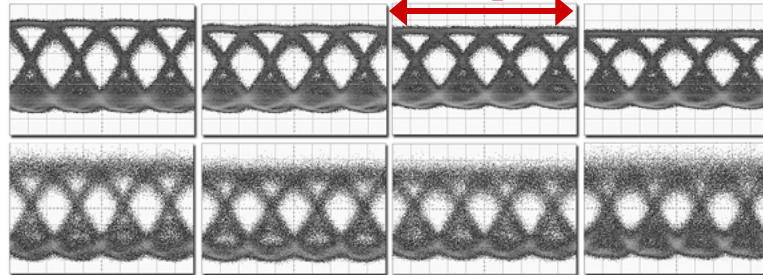


Silicon Photonic Waveguide Technology

1.28 Tb/s Data Transmission Experiment (occupies small slice of available WG BW)



[Vlasov and McNab, *Optics Express* 12 (8) 1622 (2004)]



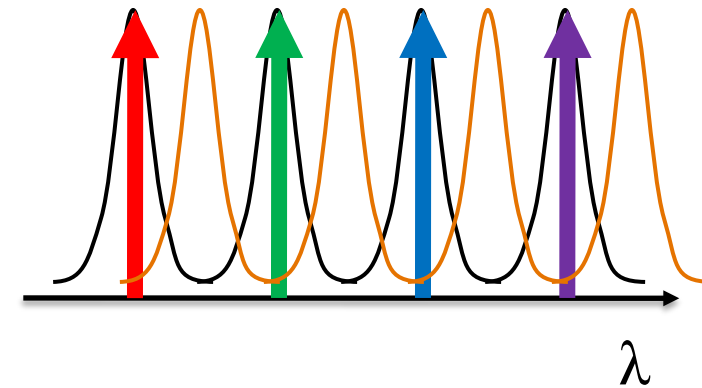
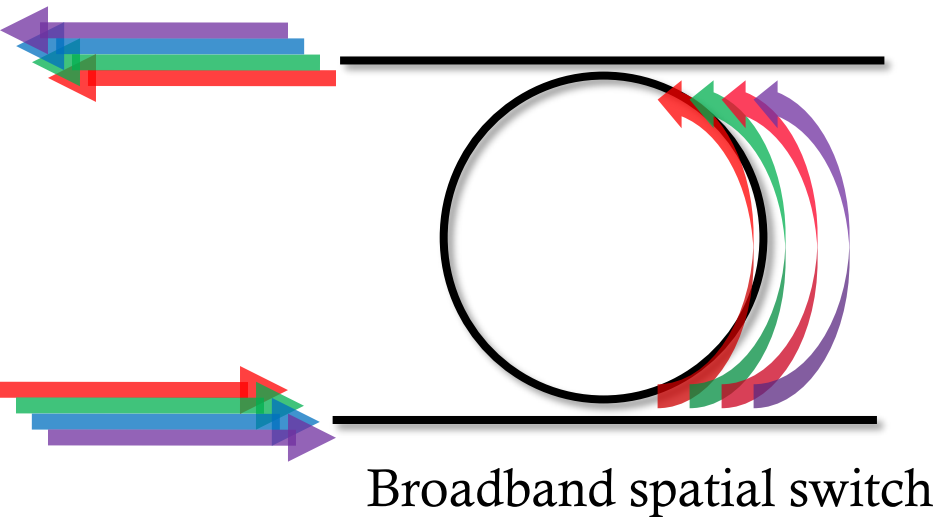
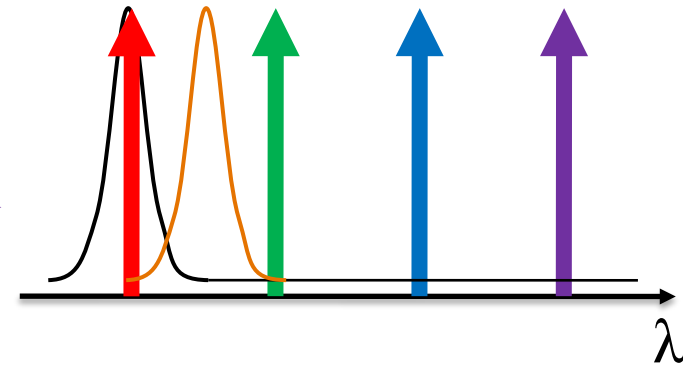
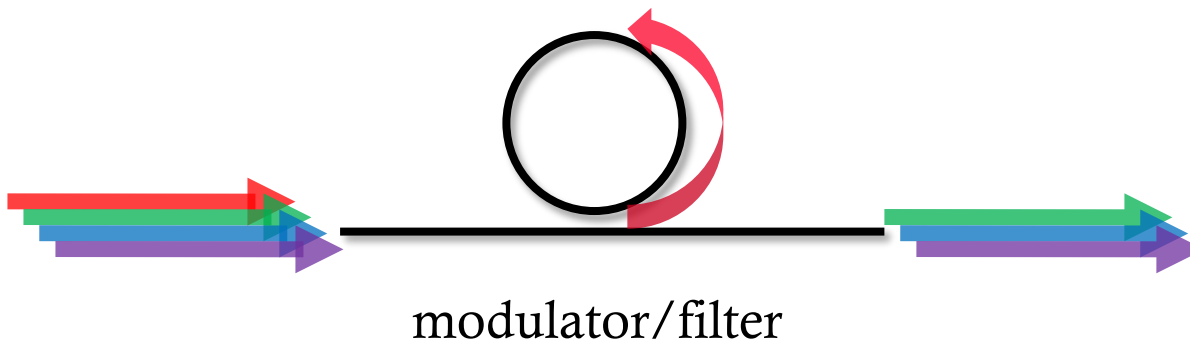
C23 **C28** **C46** **C51**
(1559 nm) **(1555 nm)** **(1541 nm)** **(1537 nm)**

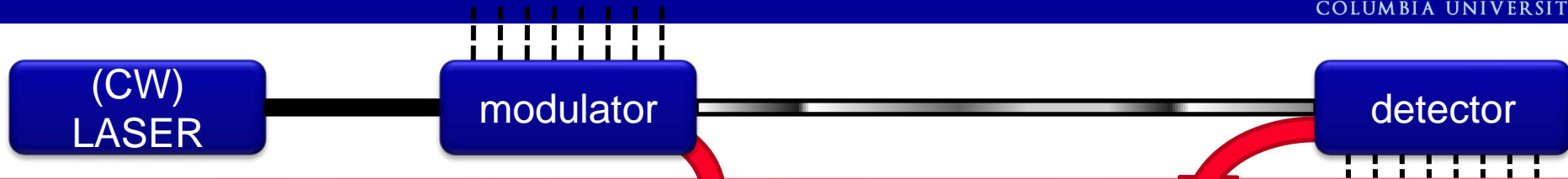
[B. G. Lee *et al.*, *Photon. Technol. Lett.* 20 (10) 767 (2008)]

Silicon photonic waveguides provide low-power optical interconnects in CMOS-compatible platform.

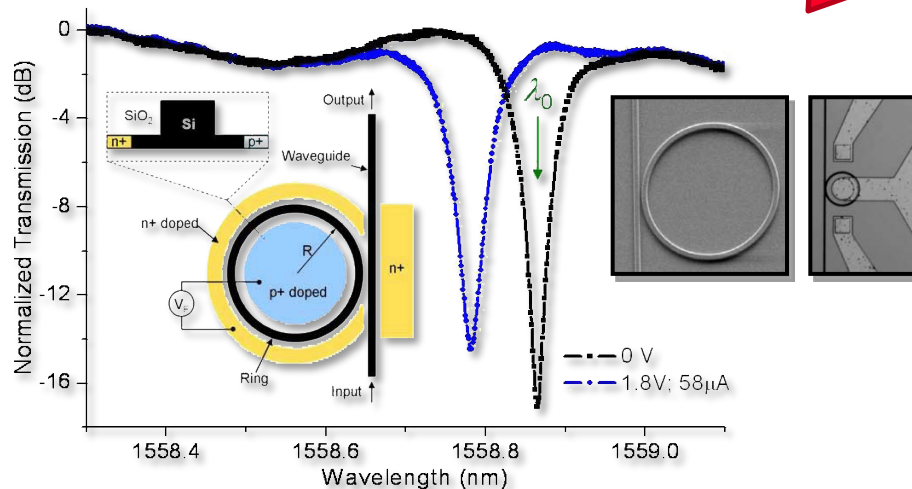
Low-loss (1.7 dB/cm), high-bandwidth (> 200 nm) silicon photonic waveguides can be fabricated in commercial CMOS process.

Ring Resonator Operation



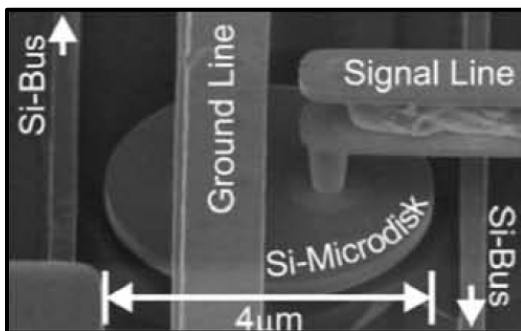


18 Gb/s demonstrated



[M Lipson, *Optics Express* (2007)]

- 85 fJ/bit demonstrated at 10 Gb/s
- Scalable to < 25 fJ/bit



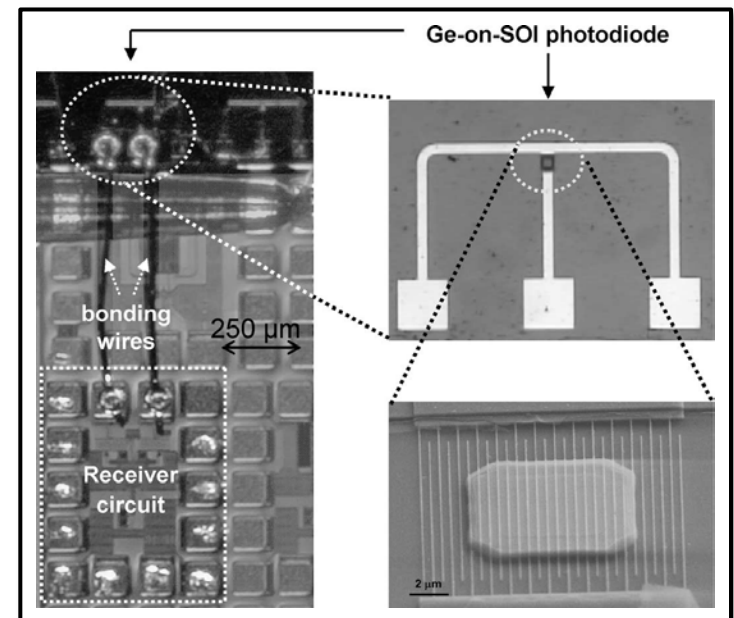
[M Watts, *Group Four Photonics* (2008)]

Ge-on-Si Detectors:

- 40-GHz bandwidths
- 1 A/W responsivities

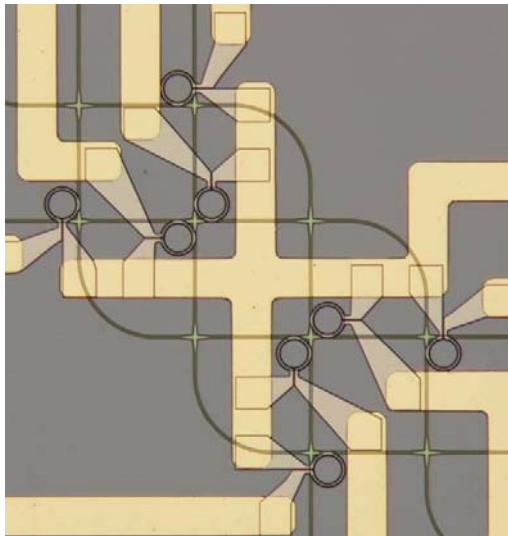
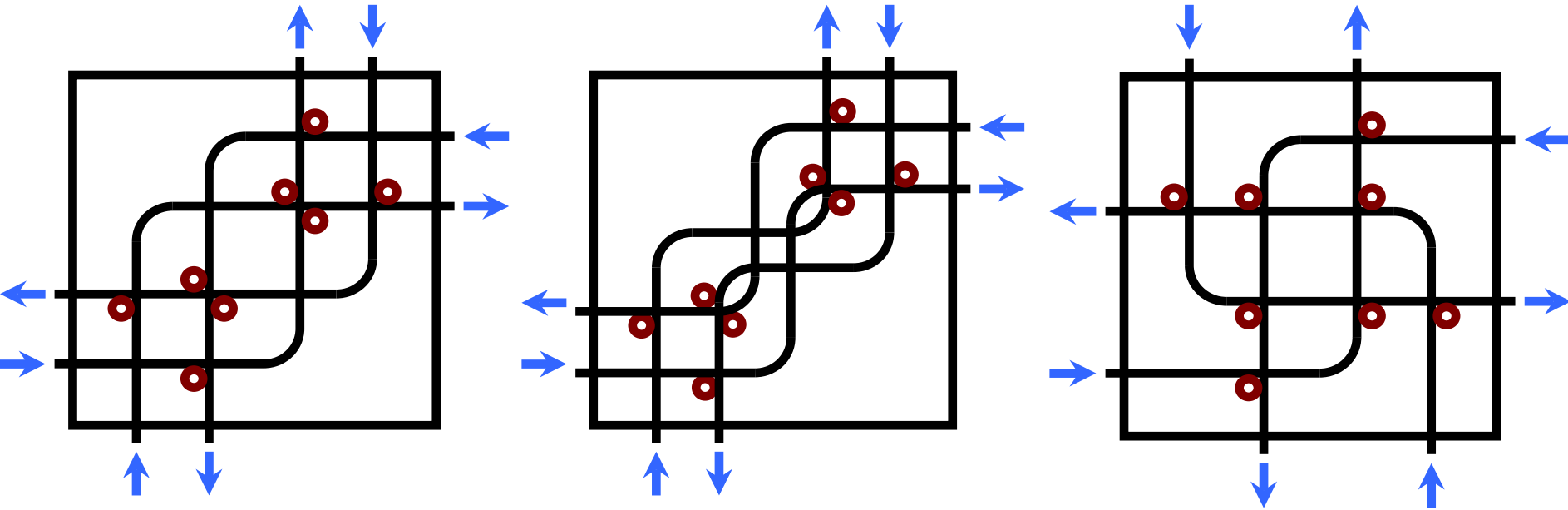
Receivers (detectors w/ CMOS amplifiers):

- 1.1 pJ/bit demonstrated at 10 Gb/s
- Scalable to < 50 fJ/bit

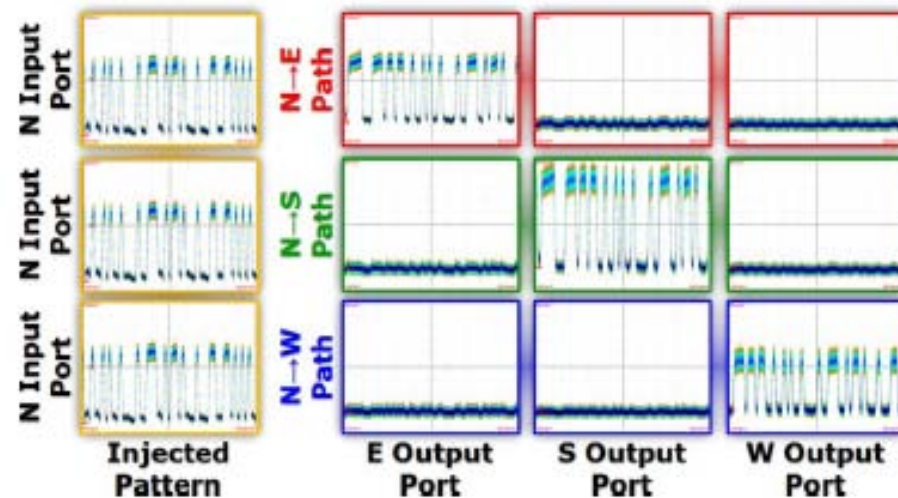


[S Koester, *J. Lightw. Technol.* (2007)]

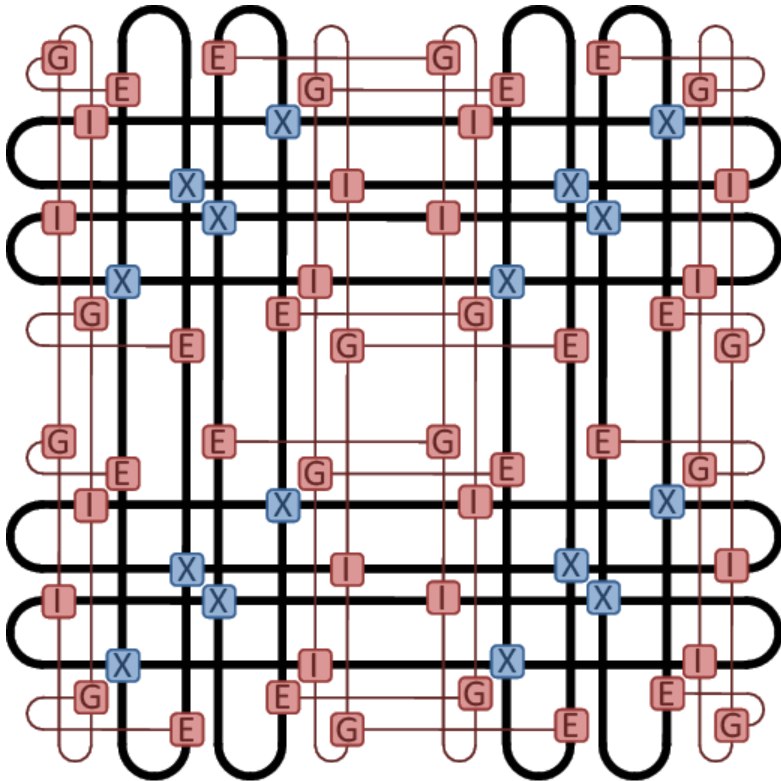
Higher Order Switch Designs



Single-Channel (1546-nm) Routing Verification for Three Switch Configurations: **N→E**, **N→S**, and **N→W**

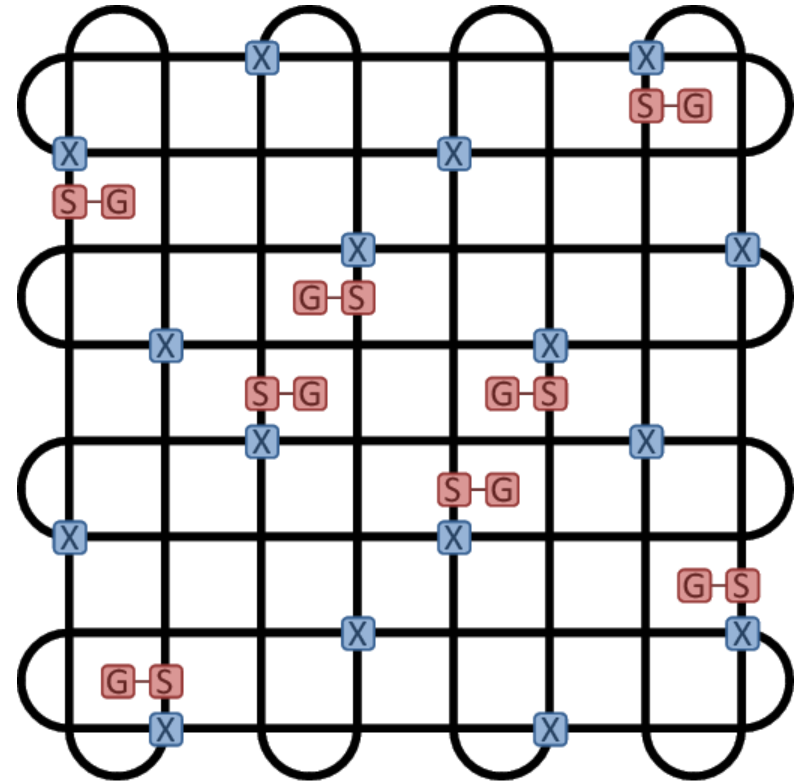


- **Photonic Torus**



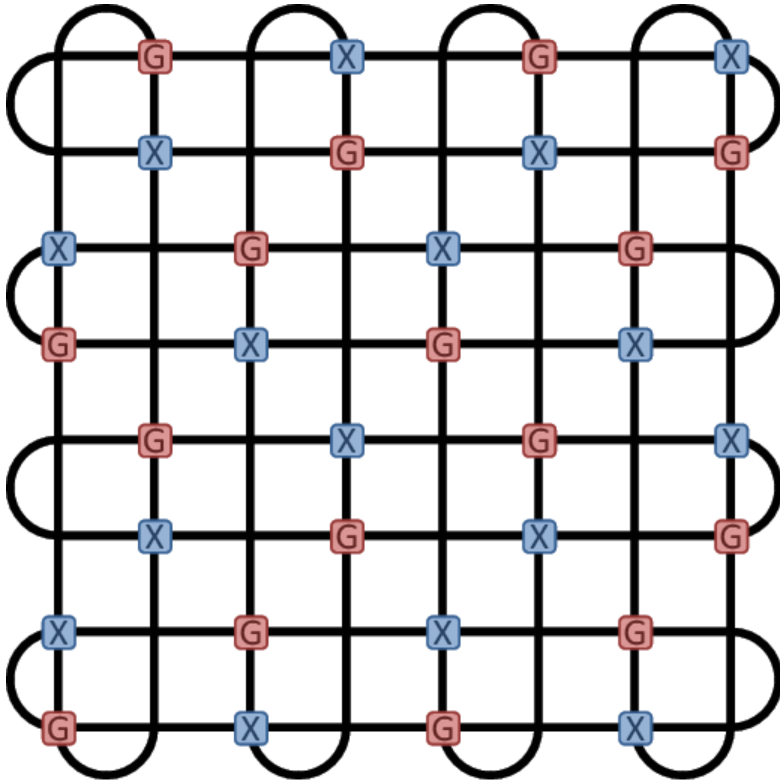
[A. Shacham *et al.*, Trans. on Comput., 2008]

- **Nonblocking Photonic Torus**

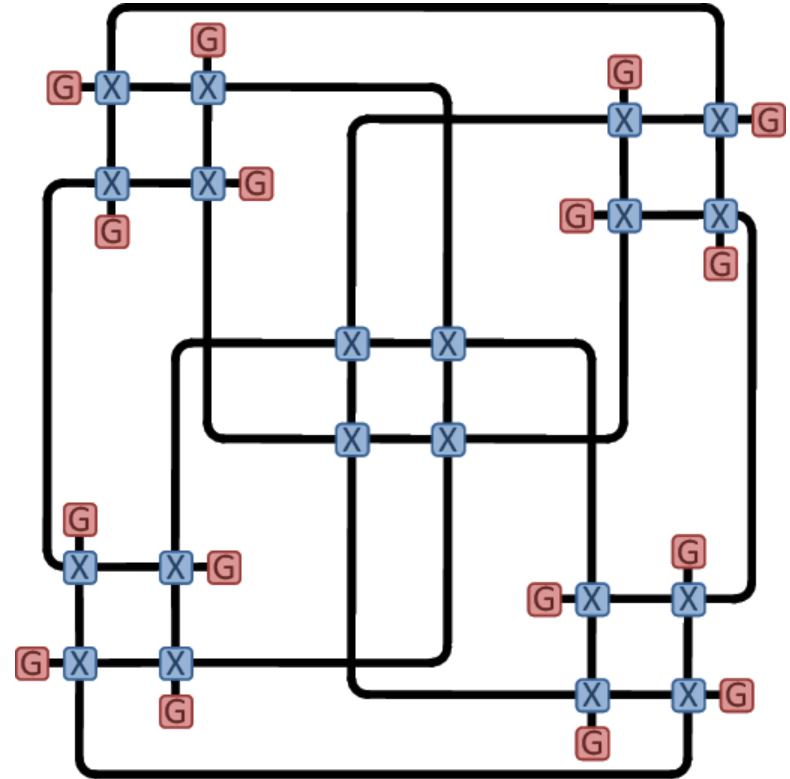


[M. Petracca *et al.* IEEE Micro, 2008]

- **TorusNX**



- **Square Root**

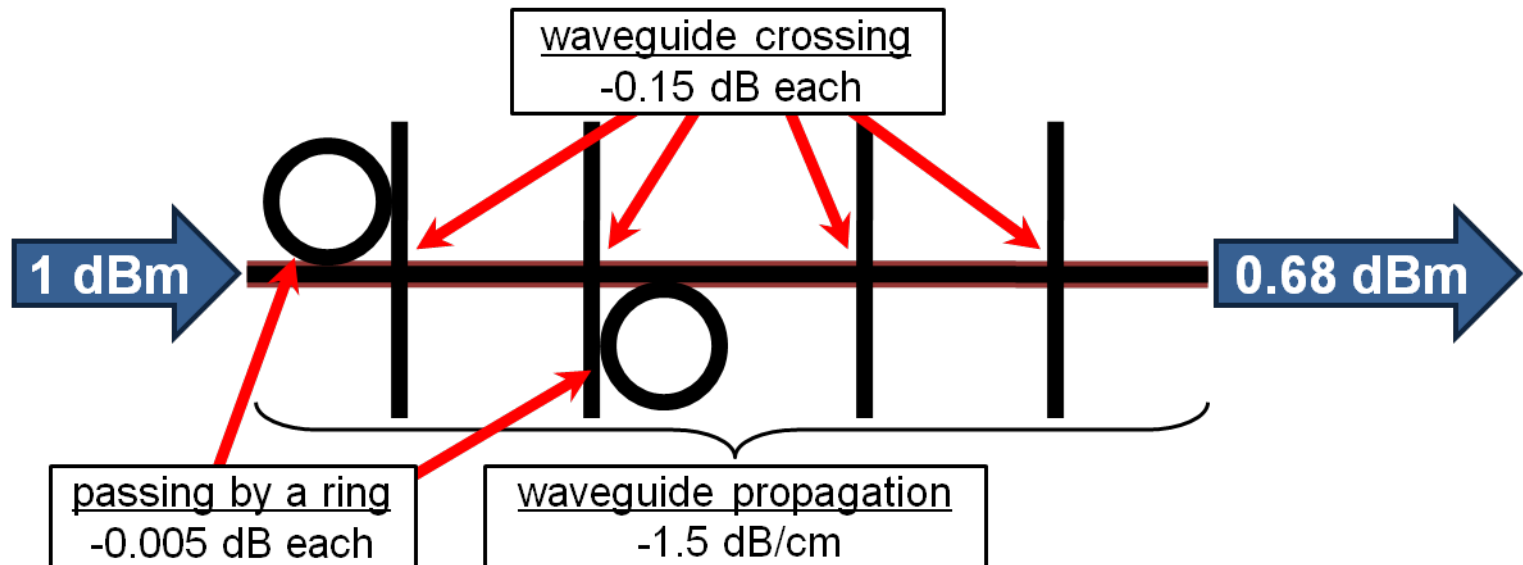
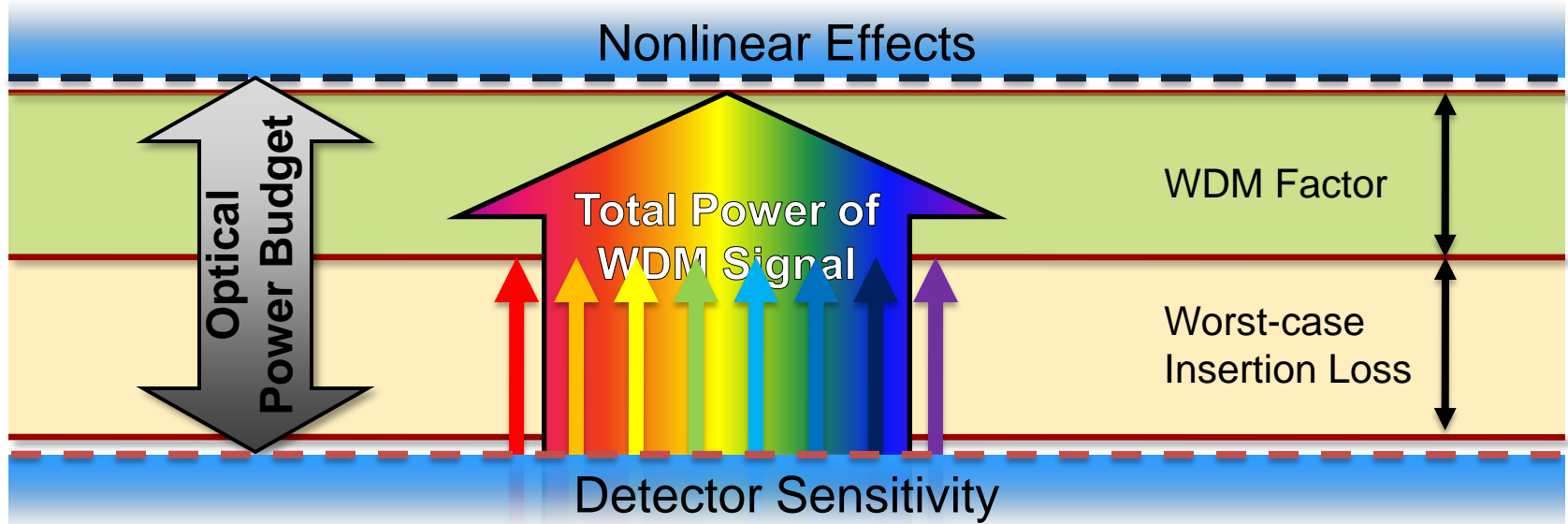


[J. Chan et al. JLT, May 2010]

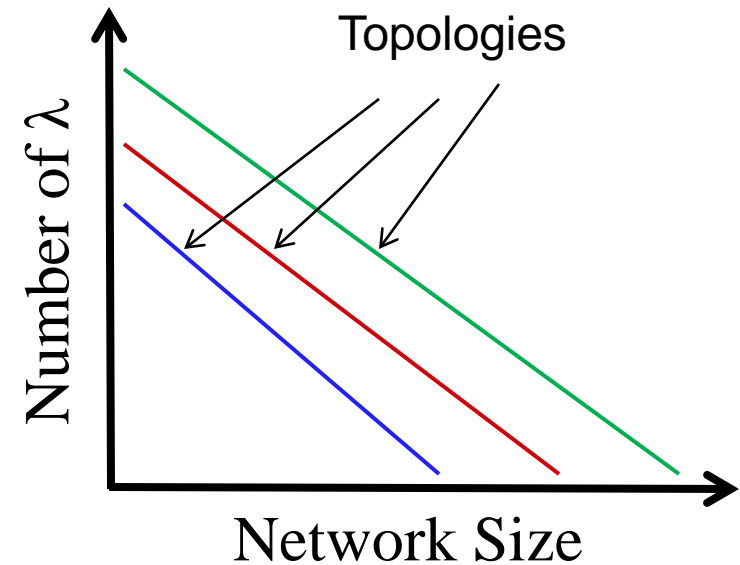
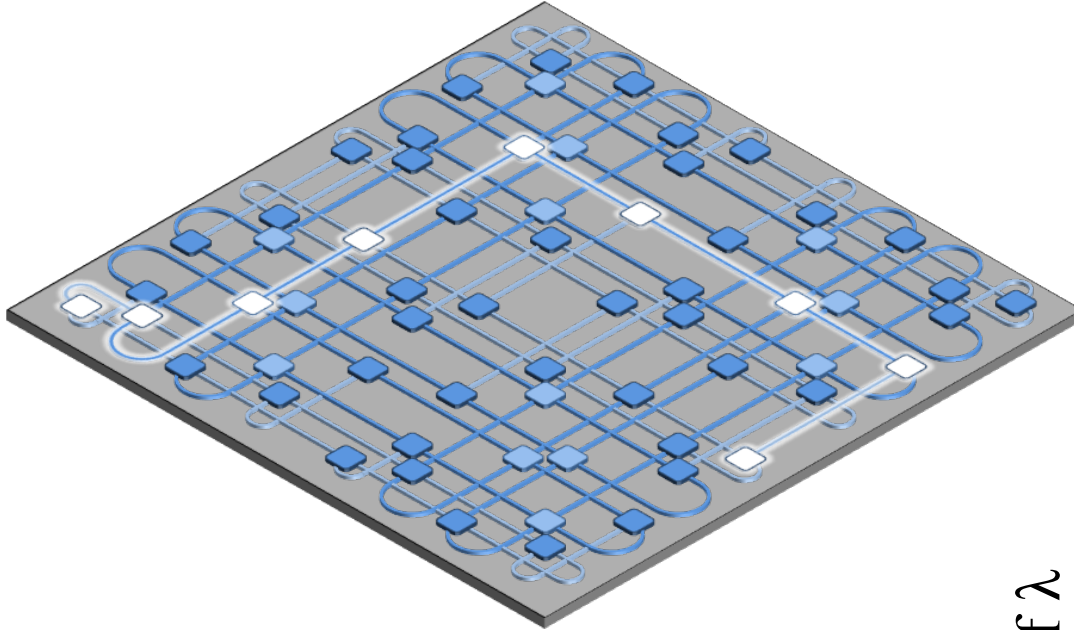
Photonic Plane Characteristics

- **Insertion Loss**
- Noise
- Power

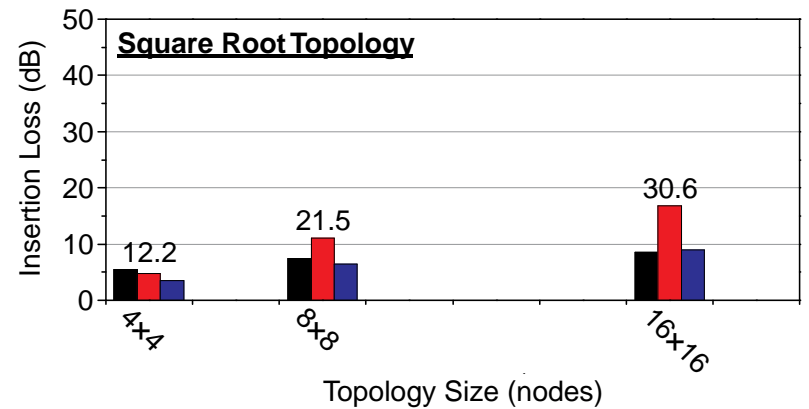
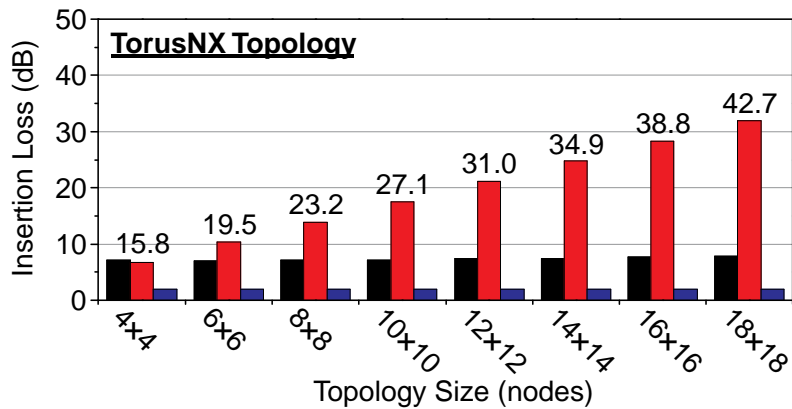
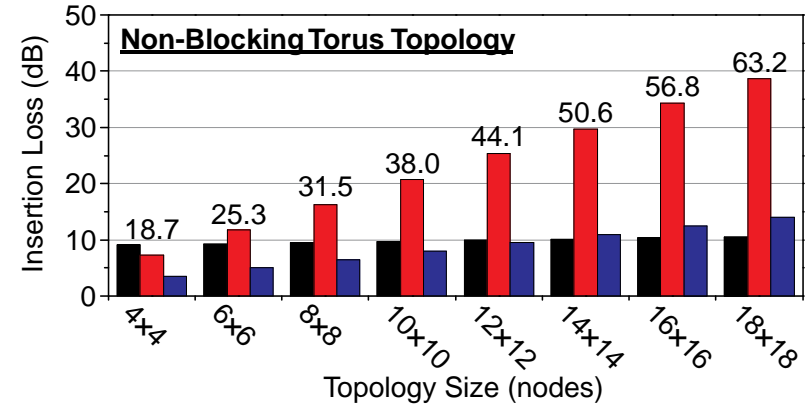
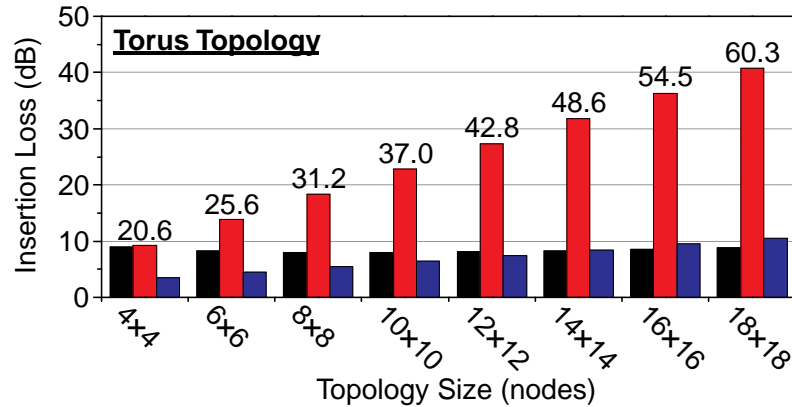
Insertion Loss and Optical Power Budget



Insertion Loss vs. Bandwidth



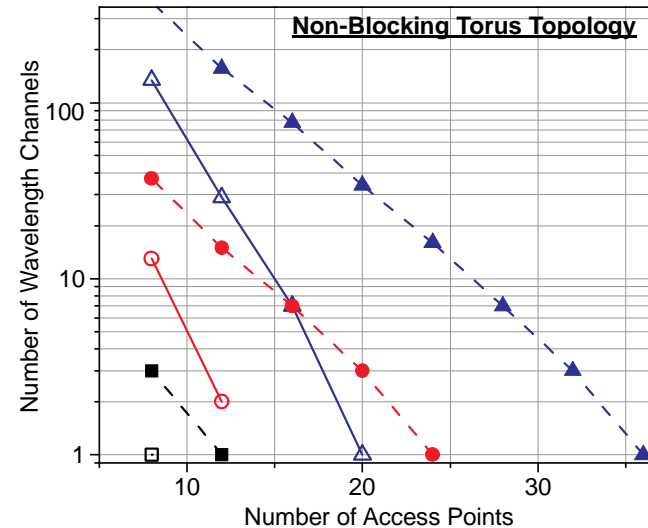
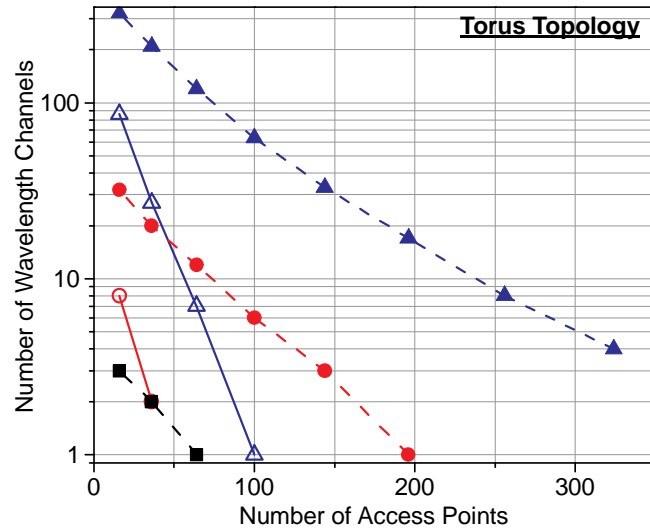
Simulation Results



■ Propagation ■ Crossing ■ Dropping Into a Ring

Simulation Results

Original is based on the IL results from previous slide, *Improved* is based on a hypothetical improvement in crossing loss from 0.15 dB to 0.05 dB.



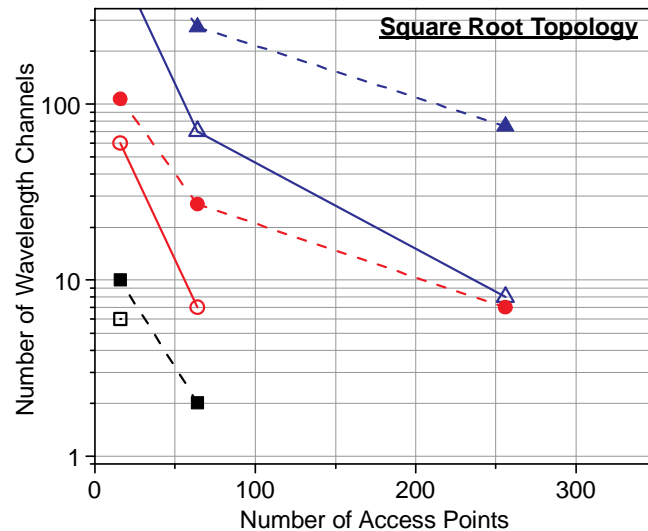
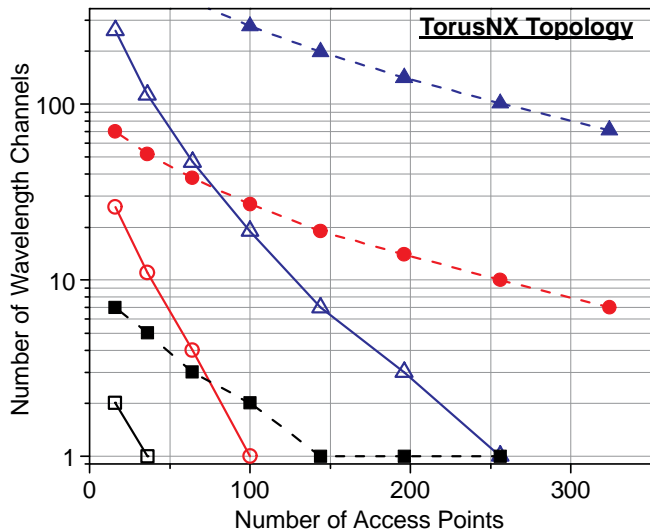
Optical power budget

Original

- 20 dB
- 30 dB
- 40 dB

Improved

- 20 dB
- 30 dB
- 40 dB



Optical power budget

Original

- 20 dB
- 30 dB
- 40 dB

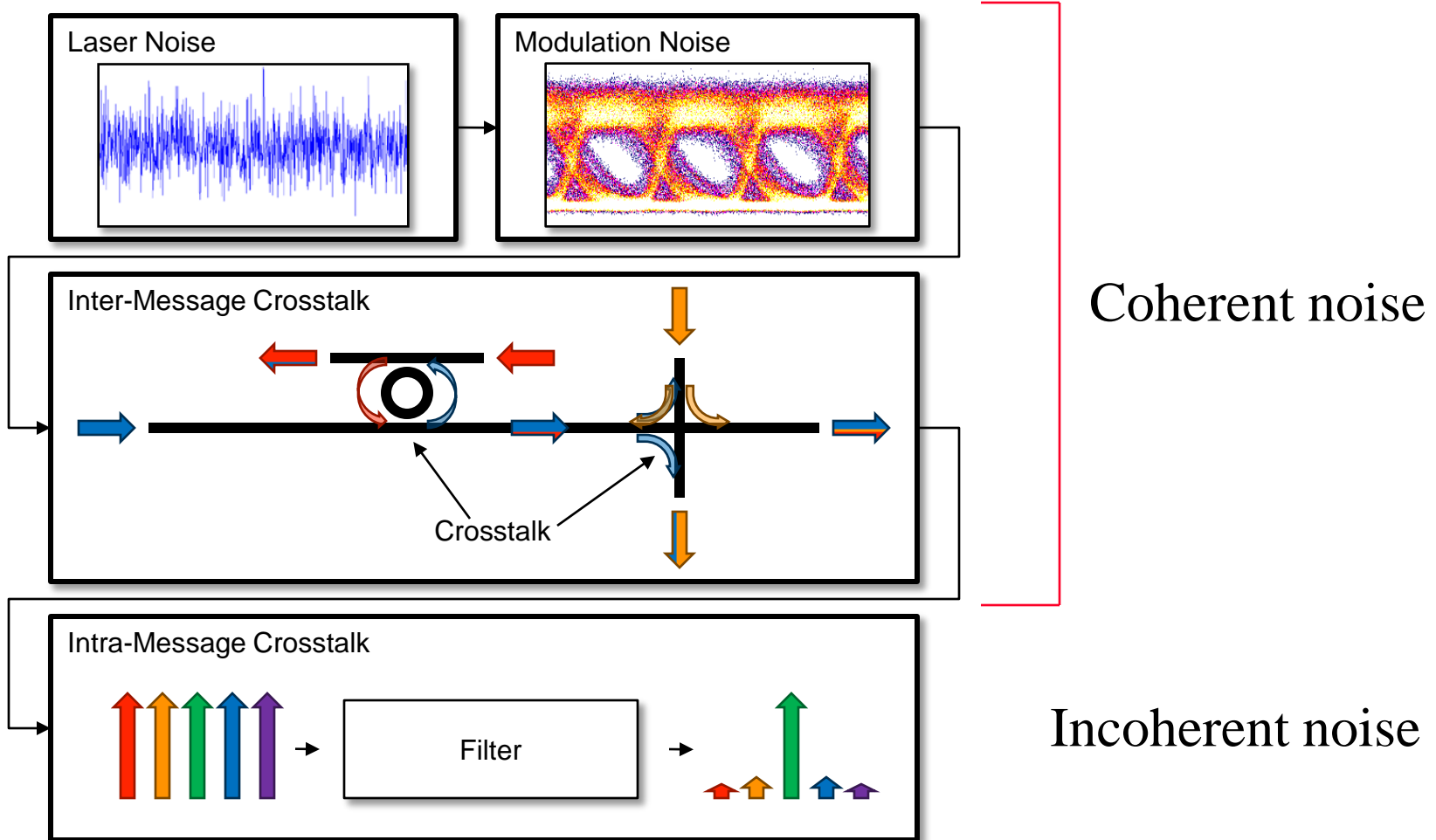
Improved

- 20 dB
- 30 dB
- 40 dB

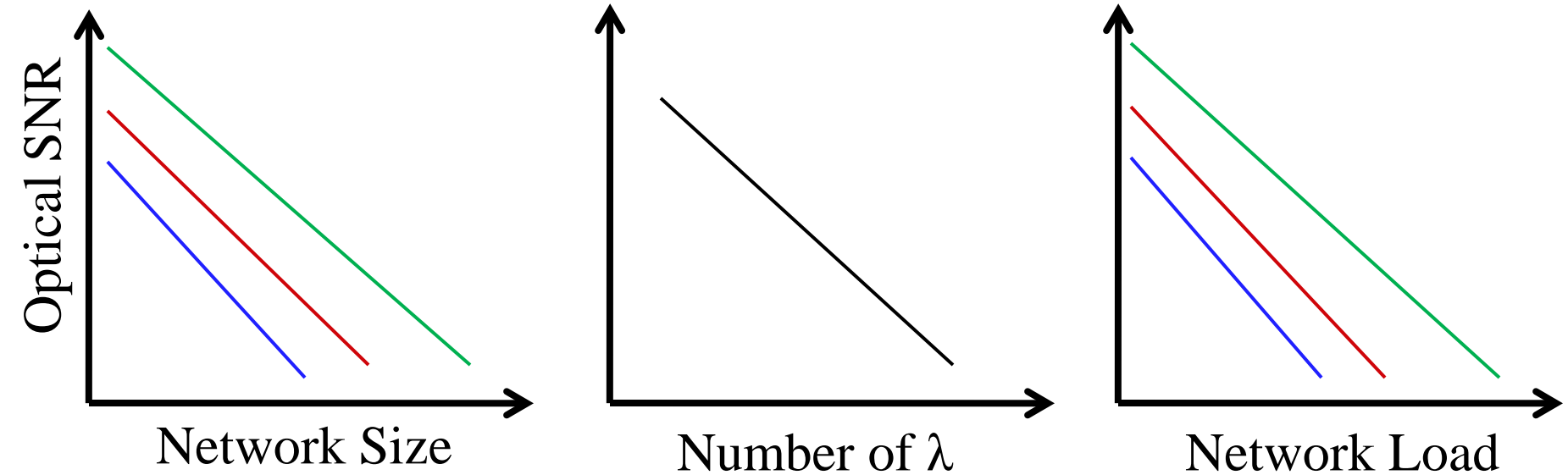
Photonic Plane Characteristics

- Insertion Loss
- **Noise**
- Power

Noise and Crosstalk



Effects of Noise

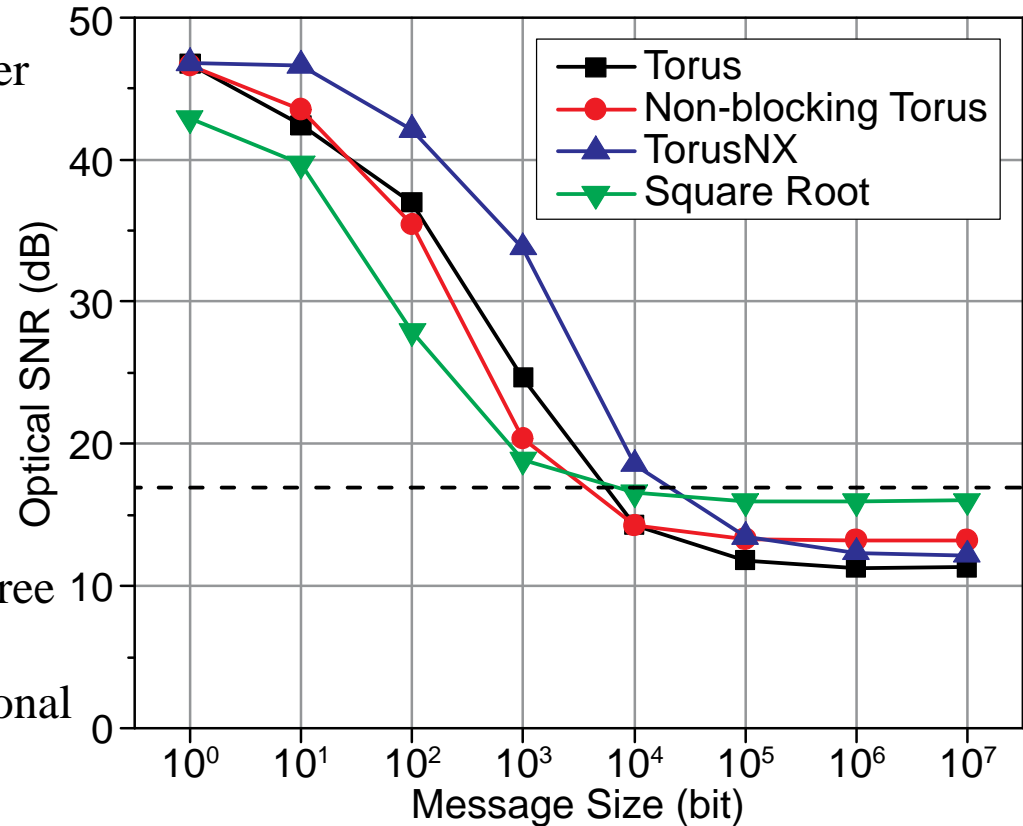


Results

- Results are plotted for network size of 8×8 at saturation, at the detectors.
- Maximum OSNR = ~ 45 dB (due to laser noise)
- Minimum OSNR < 17 dB (due to message-to-message crosstalk)
- Variations between networks due to varying likelihood of two message intersecting on network topology.

System Performance

- SNR measures the likelihood of error-free transmission.
- Lower SNR designs will require additional retransmission, resulting in lower throughput performance.

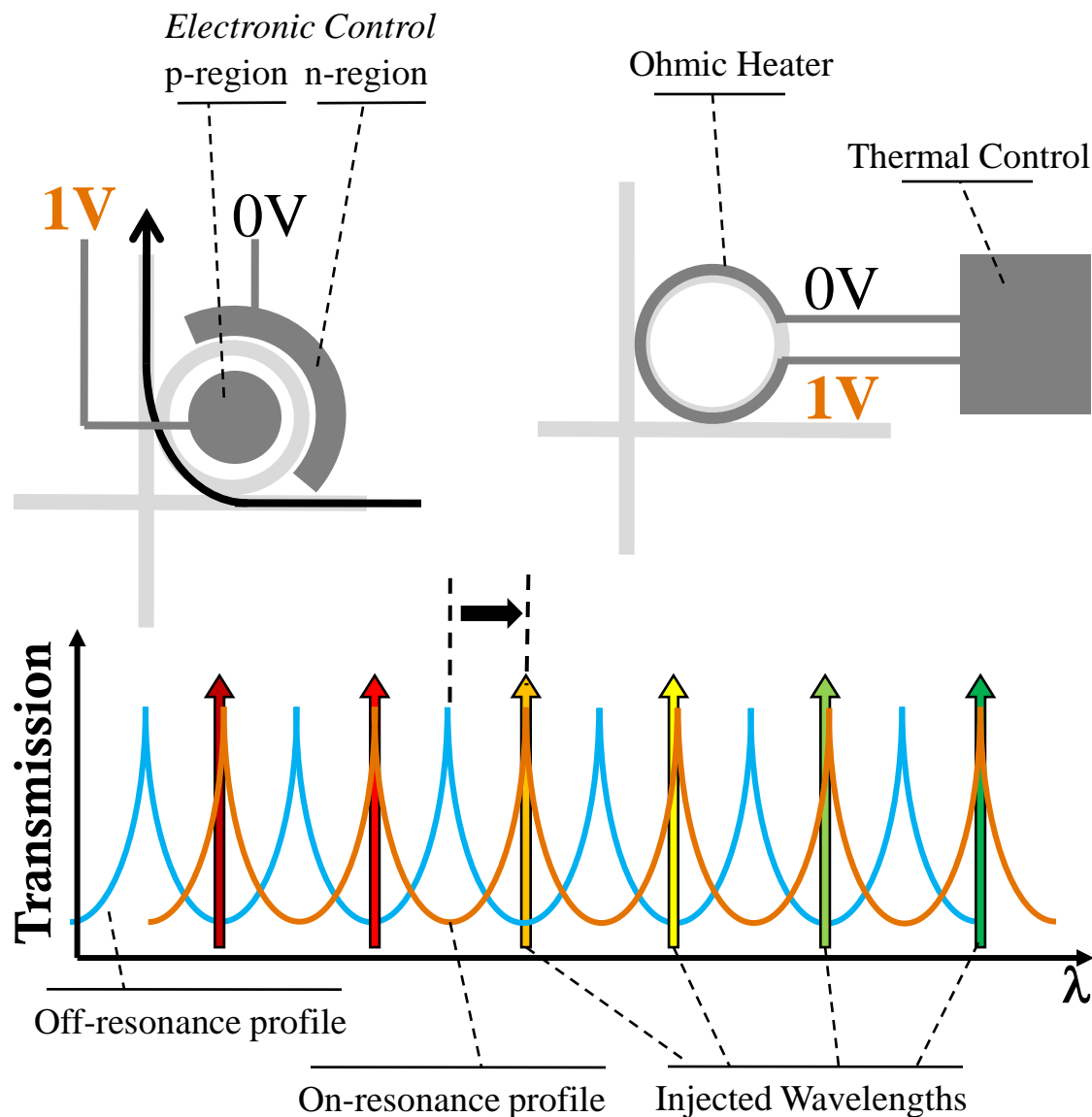


The line at OSNR=16.9 dB is where a bit-error-rate of 10^{-12} can be achieved, assuming an ideal binary receiver circuit and orthogonal signaling.

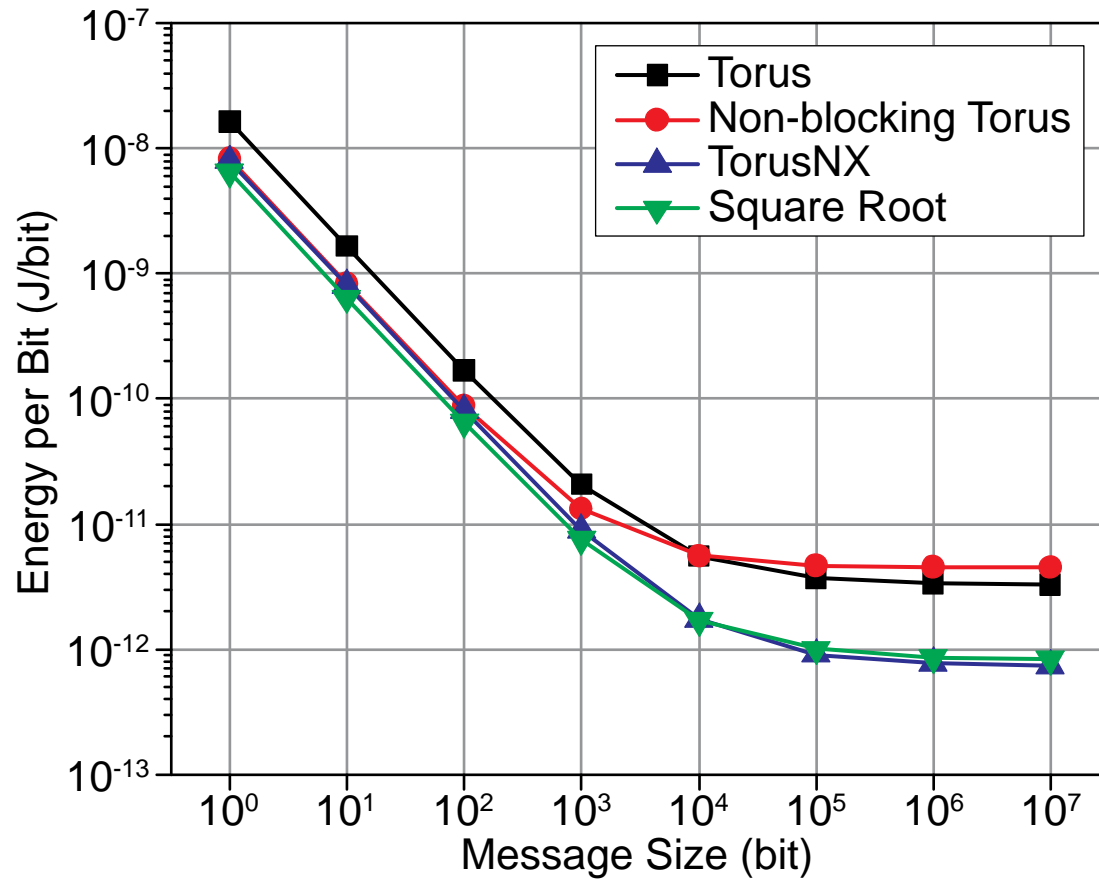
Photonic Plane Characteristics

- Insertion Loss
- Noise
- **Power**

- **Laser Power**
- **Active Power**
 - Modulating
 - Detecting
 - Broadband
- **Static Power**
 - Thermal tuning
- **Tx\Rx Power**
 - Drivers
 - TIAs

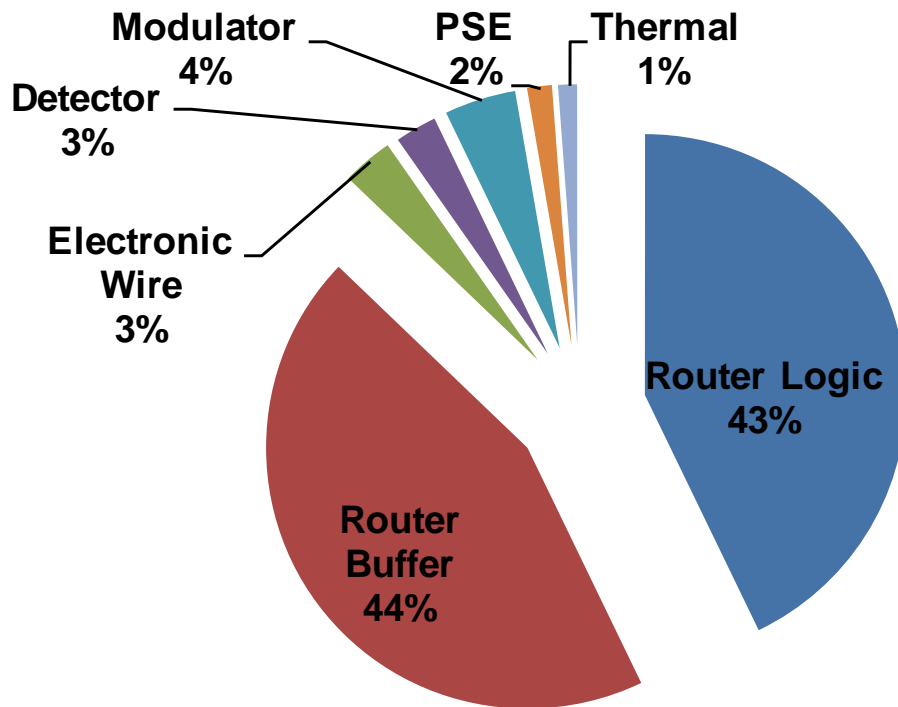


Energy Per Bit



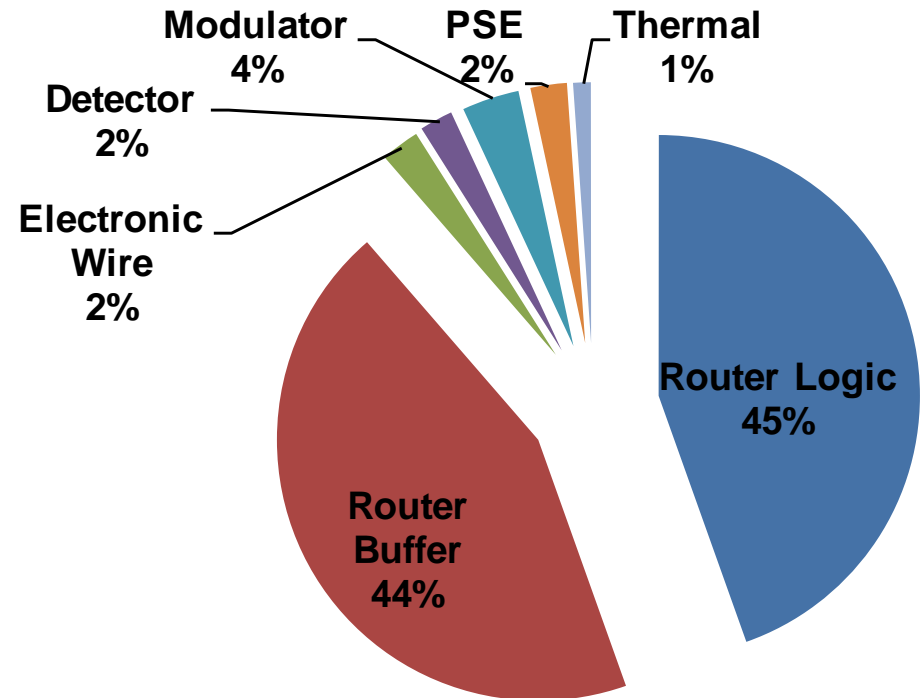
Power Breakdown

Torus Topology



- 12 wavelengths @ 10 Gbps/each
- Power Dissipation = 4.31 W

Nonblocking Torus Topology

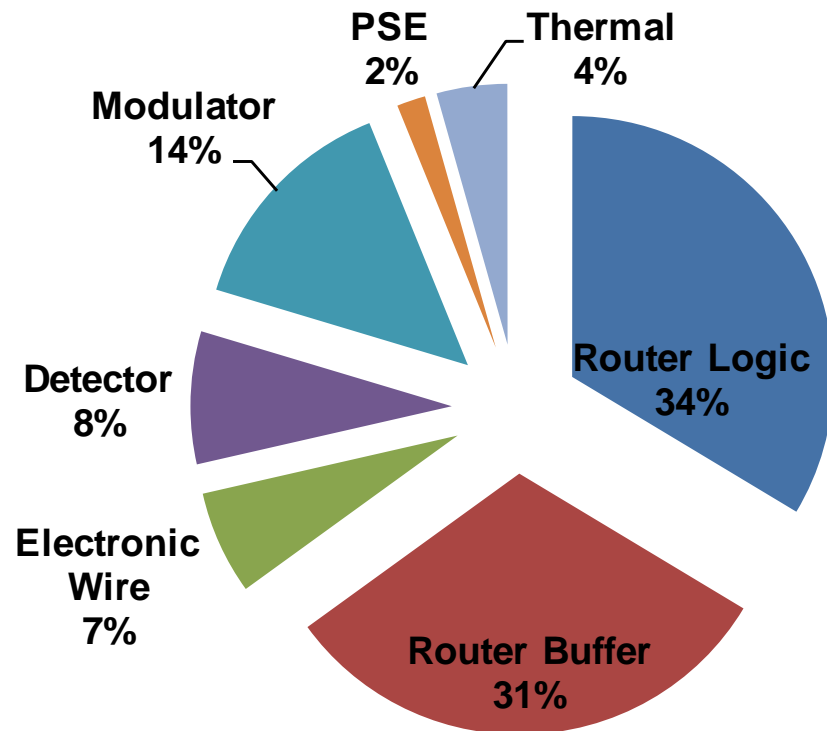


- 7 wavelengths @ 10 Gbps/each
- Power Dissipation = 1.59 W

- Results based on randomly generated traffic with message sizes of 100 kbit, with network in saturation.
- Data was collected on 64 nodes topologies constrained to a total surface area of 2 cm × 2 cm.

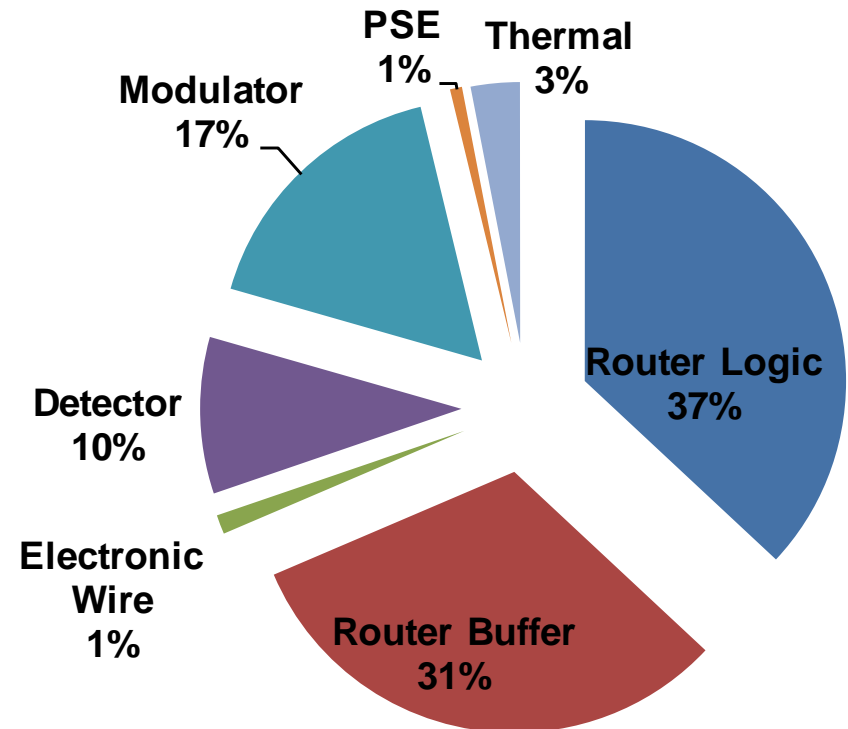
Power Breakdown

Square Root Topology



- 27 wavelengths @ 10 Gbps/each
- Power Dissipation = 1.89 W

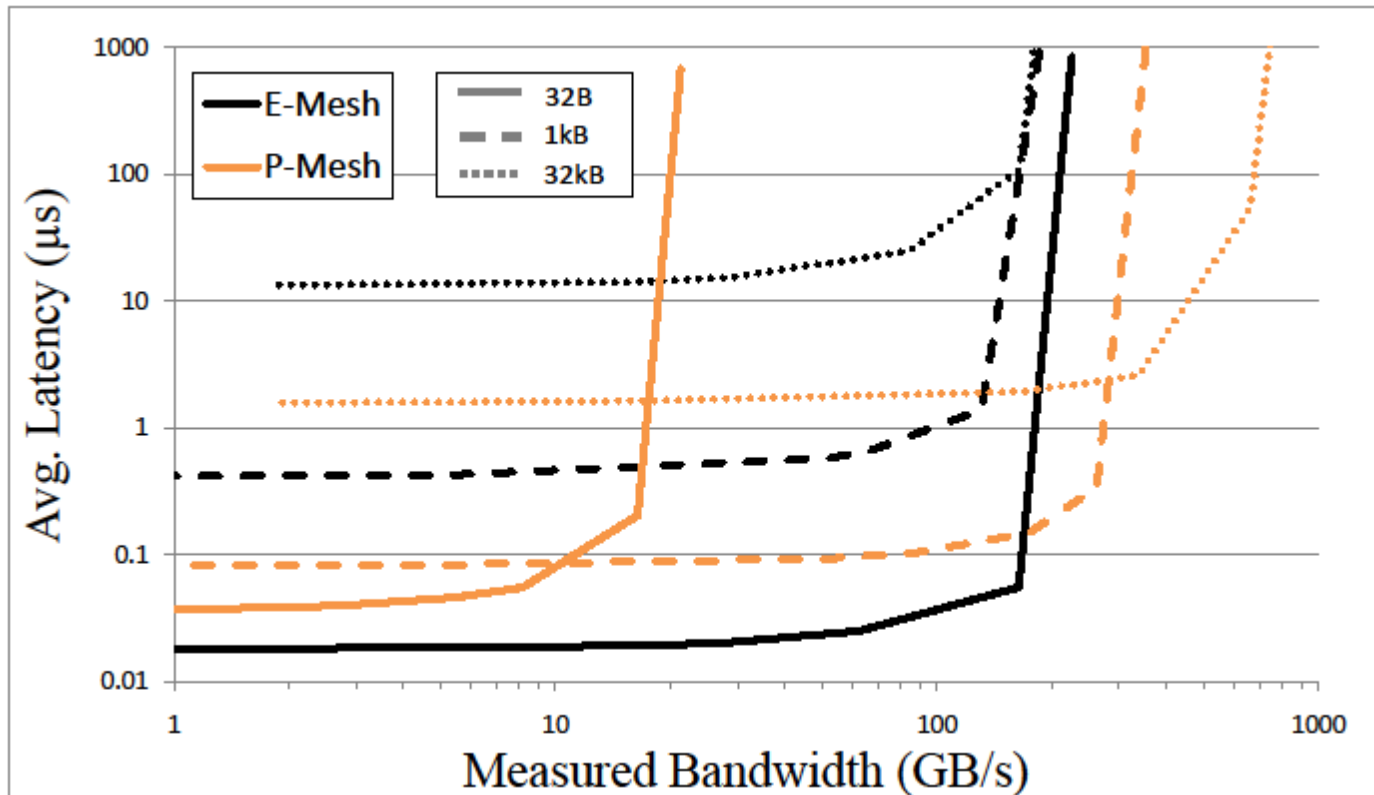
TorusNX Topology

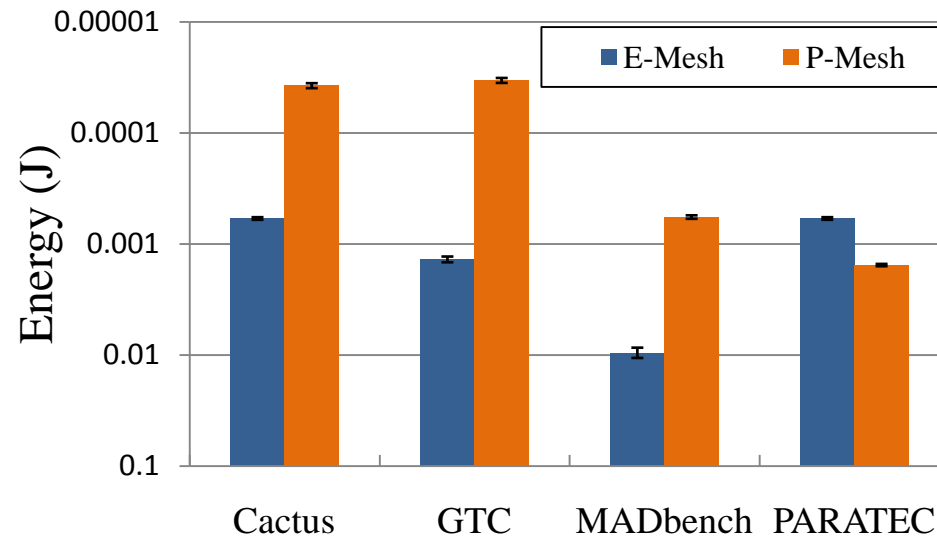
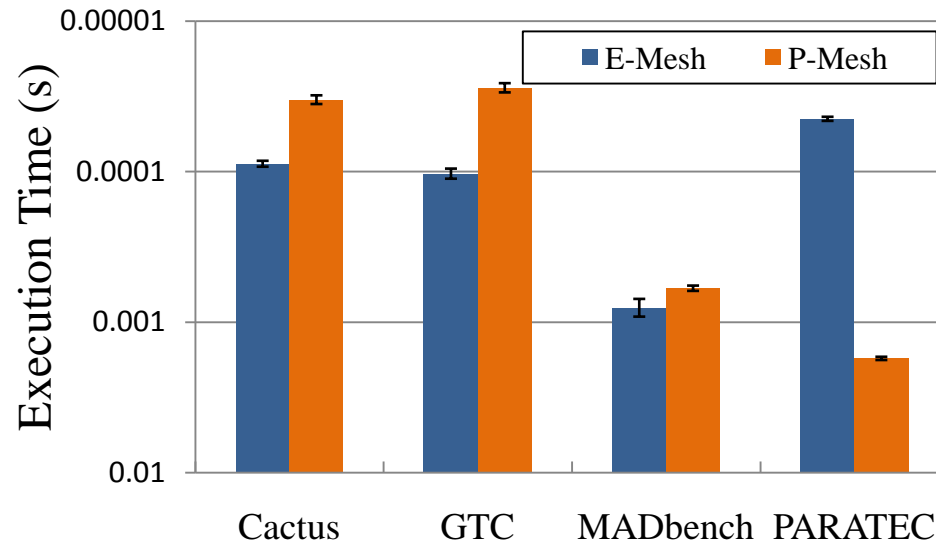


- 38 wavelengths @ 10 Gbps/each
- Power Dissipation = 3.22 W

Performance

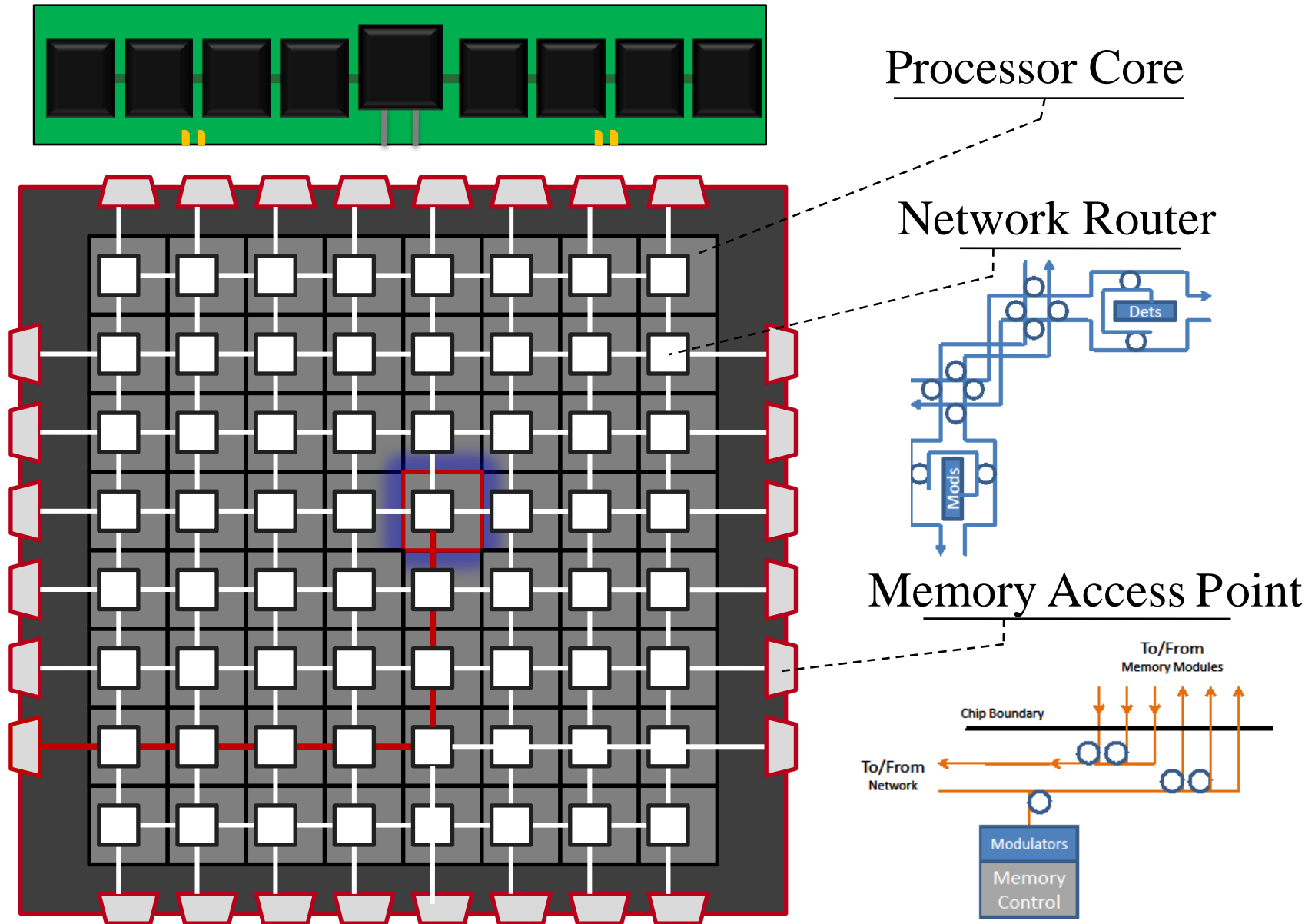
- Uniform random traffic
- 256 cores, 64-node network



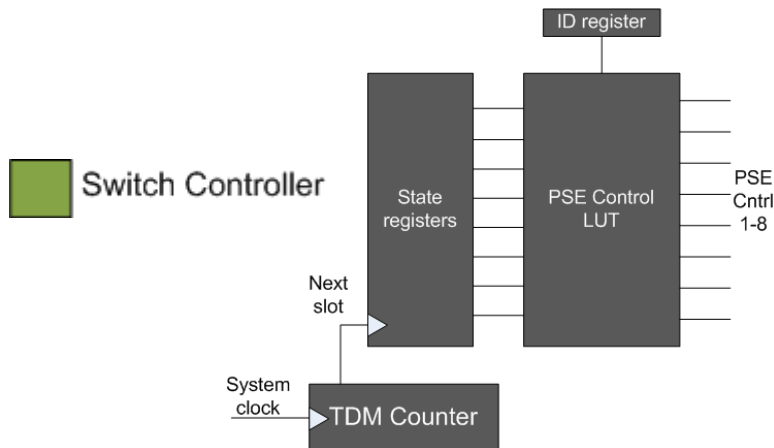
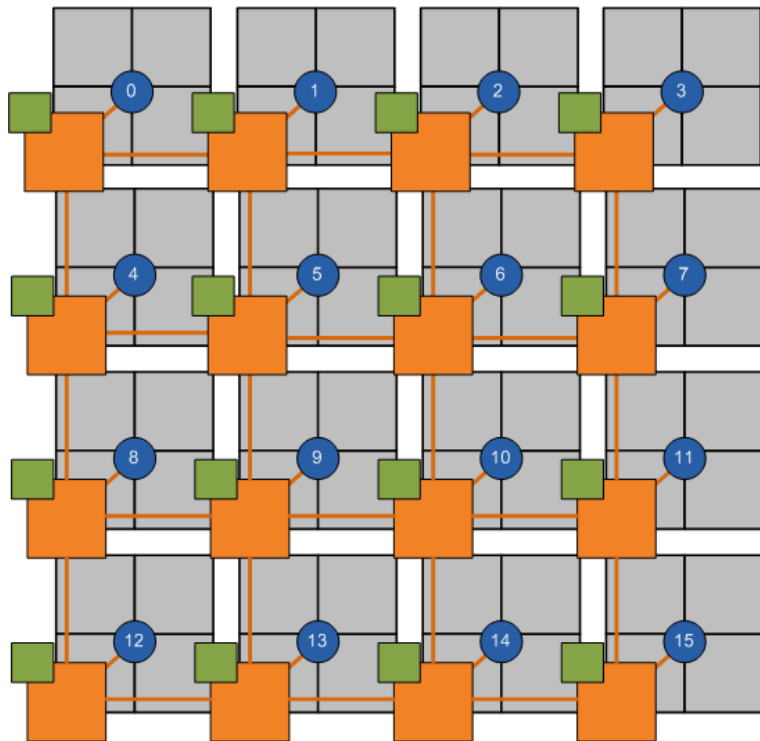


Other Interesting Issues

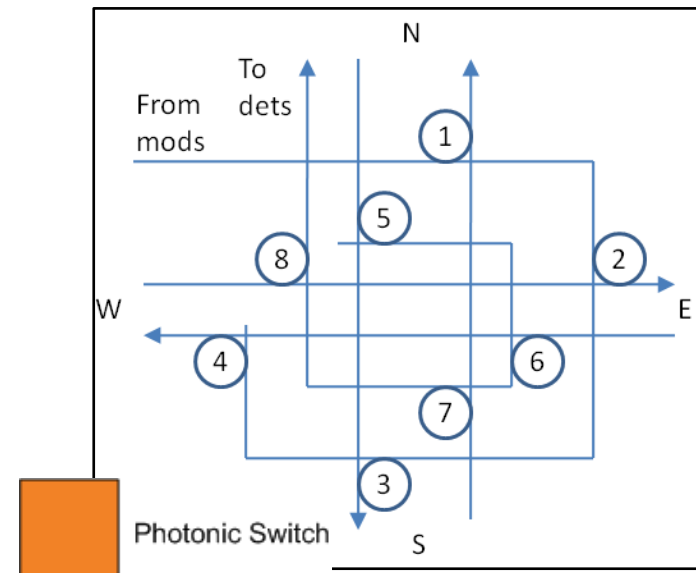
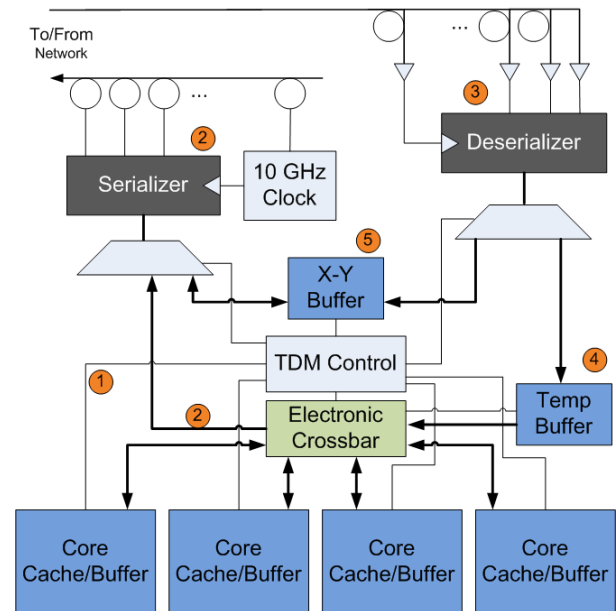
Memory Access



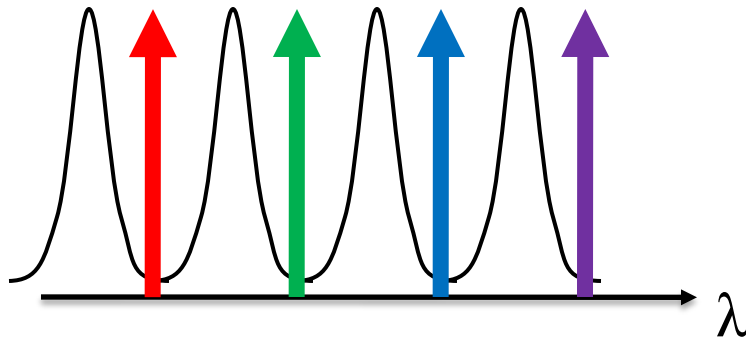
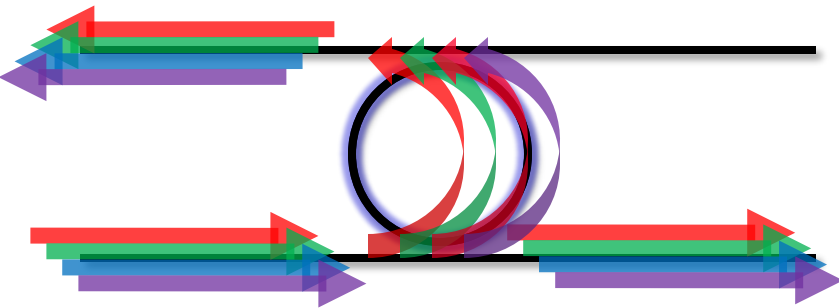
Other Arbitration Means - TDM



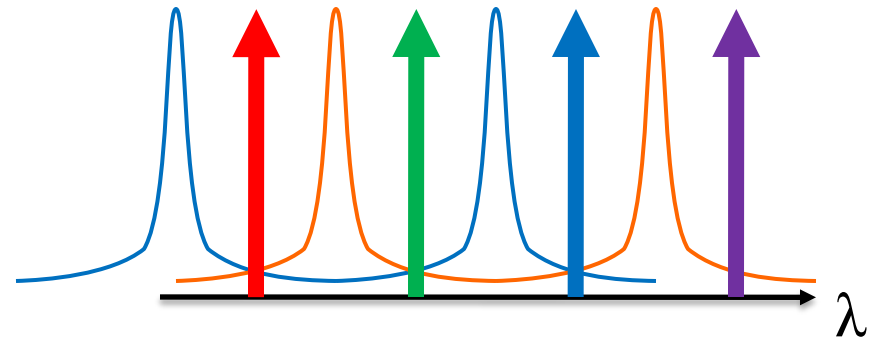
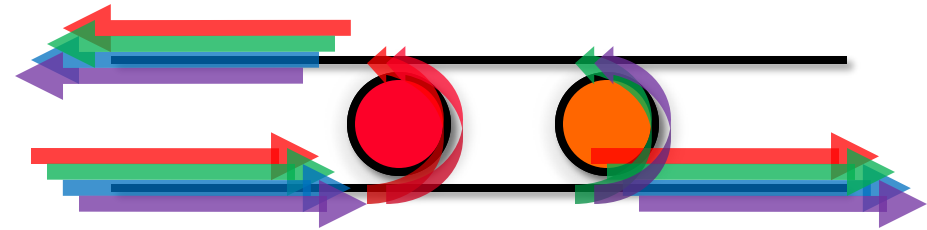
 Network Gateway



- **Original**



- **Re-design**



- Scalable number of WDM channels

- Some **applications / programming models** definitely well-suited to a circuit-switched photonic network
- Interesting tradeoffs and design space
 - Photonic physical layout / design
 - System-level benefits from device improvement
 - Network-level improvements